

OBJECTIVE BAYESIAN ESTIMATION
FOR THE NUMBER OF CLASSES IN A POPULATION
USING JEFFREYS AND REFERENCE PRIORS

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by
Kathryn Jo-Anne Barger

August 2008

© 2008 Kathryn Jo-Anne Barger

OBJECTIVE BAYESIAN ESTIMATION
FOR THE NUMBER OF CLASSES IN A POPULATION
USING JEFFREYS AND REFERENCE PRIORS

Kathryn Jo-Anne Barger, Ph.D.

Cornell University 2008

Estimation of the number of classes in a closed population is a problem that arises in many different subject areas. A common application occurs in animal populations where there is interest in determining the number of different species, also the diversity or species richness, of the population.

In this dissertation a class of models are considered, all with the common assumption that the number of individual items from each class, contributed to the sample, is a Poisson random variable. In order to conduct objective Bayesian inference, Jeffreys and reference priors are derived from the full likelihood. The Jeffreys and reference priors are functions of the information for the model parameters. The information is calculated in part using the linear difference score for integer parameter models (Lindsay & Roeder 1987).

A main accomplishment of this dissertation is deriving the form of the Jeffreys and reference priors for the number of classes. Both of the priors for the Jeffreys and reference methods factor into two independent priors, one for the parameter of interest and one for the nuisance parameters. This gives a justification for choosing these priors independent a priori, as is generally done for this problem. In this dissertation we prove that the posterior determined by these priors is proper for some low dimensional problems. With increased dimensionality, the form of the prior and posterior becomes increasingly intractable and we propose methods to deal with these difficulties. Microbial samples from the Framvaren

Fjord (Behnke et al. 2006) and Lepidoptera data (Fisher et al. 1943) are used to illustrate that these priors can be used to implement a fully Bayesian procedure for the number of classes. The reference prior is comparable to maximum likelihood results, while the Jeffreys prior deviates from the frequentist estimates in models with larger numbers of parameters.

BIOGRAPHICAL SKETCH

Kathryn Barger graduated from Carmel High School in Carmel, NY in 1999. She received a Bachelor of Music in violin performance with a second major in mathematics from the State University of New York at Potsdam in 2003. She received a Master of Science in Statistics from Cornell University in 2006.

ACKNOWLEDGMENTS

I would like to thank the Department of Statistical Science at Cornell University for their financial support. I would also like to thank Dr. Thorsten Stoeck and the Stoeck lab, Technische Universität Kaiserslautern, Germany, for the use of their data set from the Framvaren Fjord (Behnke et al. 2006). I would like to thank the members of my thesis committee, Dr. Giles Hooker, Dr. Martin Wells and Dr. John Bunge, for dedicating their time in helping me complete my dissertation. Finally, I sincerely thank my thesis advisor, Dr. John Bunge, for his unconditional support in the completion of my graduate education.

TABLE OF CONTENTS

Biographical Sketch	iii
Acknowledgments	iv
Table of Contents	v
List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Estimating the Number of Classes	3
1.2 The Statistical Model	6
1.2.1 A Poisson Model	6
1.2.2 The Likelihood	8
1.2.3 Information	10
1.3 An Objective Bayesian Approach	15
1.3.1 Bayesian Estimation for the Species Problem	15
1.3.2 Objective Priors	18
1.4 Data Sets	21
1.4.1 Framvaren Fjord	21
1.4.2 Lepidoptera	22
1.5 Layout of the Dissertation	23
2 Estimation Conditional on the Nuisance Parameters	26
2.1 Jeffreys and Reference Priors	26
2.2 A Class of Priors for the Binomial Likelihood	28
3 One Nuisance Parameter	32
3.1 Jeffreys Prior	33
3.1.1 Poisson Model	34
3.1.2 Geometric Model	38
3.2 Reference Prior	40
3.2.1 Poisson Model	41
3.2.2 Geometric Model	43
3.3 Data Analysis	45

4	Two Nuisance Parameters	52
4.1	Jeffreys Prior	52
4.2	Reference Prior	54
4.3	Negative Binomial Model	56
4.3.1	Jeffreys Prior	59
4.3.2	Reference Prior	60
4.4	Data Analysis	60
4.4.1	Framvaren Fjord	63
4.4.2	Lepidoptera	66
5	Three or More Nuisance Parameters	70
5.1	Jeffreys Prior	70
5.2	Reference Prior	72
5.3	A Three Parameter Mixture Model	75
5.4	Data Analysis	75
5.4.1	Framvaren Fjord	79
5.4.2	Lepidoptera	82
6	Summary and Conclusions	86
6.1	Comparison of Jeffreys Prior and Reference Prior	87
6.2	Comments on Bayesian, Frequentist, and Nonparametric Approaches	88
6.3	Model Selection	88
6.4	Use of the Method in Practice	90
Appendix		90
6.5	Jeffreys Prior for the Poisson Model	93
6.6	Reference Prior for the Poisson Model	98
6.7	Jeffreys Prior for the Geometric Model	104
6.8	Reference Prior for the Geometric Model	110
6.9	Jeffreys Prior for the Negative Binomial Model	116
6.10	Reference Prior for the Negative Binomial Model	124
6.11	Jeffreys Prior for the Two-mixed Geometric Model	132
6.12	Reference Prior for the Two-mixed Geometric Model	140
References		148

LIST OF FIGURES

1.1	Plot of frequency counts for the Framvaren Fjord data.	22
1.2	Plot of frequency counts for the Lepidoptera data.	24
3.1	Posterior samples of C for Poisson and geometric models	48
3.2	Diagnostic plots for models using Jeffreys priors.	49
3.3	Diagnostic plots for models using reference priors.	50
3.4	Expected frequencies for Poisson and geometric models using reference priors.	51
4.1	Posterior samples of C for the negative binomial model. (Framvaren Fjord)	64
4.2	Diagnostic plots for the negative binomial model. (Framvaren Fjord)	65
4.3	Expected frequencies for the negative binomial model. (Framvaren Fjord)	66
4.4	Posterior samples of C for the negative binomial model. (Lepidoptera)	67
4.5	Diagnostic plots for the negative binomial model. (Lepidoptera)	68
4.6	Expected frequencies for the negative binomial model. (Lepidoptera)	69
5.1	Posterior samples of C for the two-mixed geometric model. (Framvaren Fjord)	80
5.2	Diagnostic plots for the two-mixed geometric model. (Framvaren Fjord)	81
5.3	Expected frequencies for the two-mixed geometric model. (Framvaren Fjord)	82
5.4	Posterior samples of C for the two-mixed geometric model using the reference prior. (Lepidoptera)	83
5.5	Diagnostic plots for the two-mixed geometric model using the reference prior. (Lepidoptera)	84
5.6	Expected frequencies for the two-mixed geometric model using the reference prior. (Lepidoptera)	85

LIST OF TABLES

1.1	Frequency counts for Framvaren Fjord	23
1.2	Abundance counts for species of Lepidoptera	23
1.3	Frequency counts for species of Lepidoptera	24
3.1	Estimates for the Poisson model.	48
3.2	Estimates for the geometric model.	49
4.1	Estimates for the negative binomial model. (Framvaren Fjord) . .	64
4.2	Estimates for the negative binomial model. (Lepidoptera)	67
5.1	Estimates for the two-mixed geometric model. (Framvaren Fjord)	80
5.2	Estimates for the two-mixed geometric model. (Lepidoptera) . . .	84
6.1	Deviances for Framvaren Fjord data at $\tau = 20$	89
6.2	Deviances for Framvaren Fjord data at $\tau = 165$	90

Chapter 1

Introduction

An interesting application of estimating the number of classes occurs in microbiology. The diversity of microorganisms is an understudied yet important topic. Estimating the diversity of these unexplored communities is needed in the study of all biodiversity.

[M]icrobial richness predictions. . .serve as a basis for all of the paradigms of biodiversity, its role, function, and meaning. It is therefore of principal interest to know the true extent of microbial diversity, starting from that in a single environmental sample. The question therefore is: what is the total number of microbial species in a sample, habitat, and biosphere? (Hong et al. 2006)

The search for new life and study of unusual habitats on Earth has led microbiologists to study extreme environments through study of the organisms that dwell there. Microbial communities dominate these extreme environments which are often thought to be uninhabitable due to their unusual characteristics. For instance, an extreme environment may be characterized by its lack of oxygen or its high temperature. Some extreme environments are similar to prehistoric conditions on Earth, and the study of these environments hold keys to understanding the evolution of life on Earth.

The May 2008 article “Abundance and Diversity of Microbial Life in Ocean Crust” by Santelli et al., appearing in *Nature*, is a study of one of these extreme environments. Basalt covered ocean crust is known to harbor bacteria and other microorganisms. This article discusses the number of different kinds of bacteria found in the basalt habitat as well as a comparison of this habit to other marine communities. The species richness of the basalt habitat is found to be quite diverse in comparison to other marine and deep-sea environments. The authors hypothesize a complex relationship between this habitat to surrounding environments, as well as the possibility that this community is an energy source for heterotrophic microorganisms.

Microbial diversity has been studied in other extreme environments such as anoxic waters (Zuendorf et al. 2006; Jeon et al. 2006), anoxic and supersulfidic waters (Behnke et al. 2006), deep subsurface areas (Kormas et al. 2003), hypersalinic waters (Ley et al. 2006), a deep-sea hydrothermal vent (Kormas et al. 2006), nutrient limited waters (Venter et al. 2004), a hydrothermal ridge flank (Huber et al. 2003), and a volcanic hydrothermal vent habitat (Huber et al. 2006).

The motivating example given above introduces what has been called the *species problem*; that is, estimating the number of species (or type) in a population. The structure of this problem applies to many other fields outside of biology, and we name the general problem estimating the number of classes. This chapter introduces the problem of estimating the number of classes and describes other applications to the species problem. A model is described for the sampling process, and the likelihood function from the model is derived. Bayesian approaches to this problem are reviewed and an introduction on objective priors is given, including Jeffreys and reference rules. At the end of this chapter, the data sets

used in the analysis sections are introduced.

1.1 Estimating the Number of Classes

This section presents examples from various fields, such as biology, numismatics, and linguistics, which can all be viewed as applications to estimating the number of classes.

In biology, the estimation of species richness can be applied to all animals and organisms. Species richness is used to describe the number of species which reside in a certain biosphere or belong to a particular population. Knowing the number of species can help determine the complexity of an ecosystem. Finding a high level of species richness can help identify undersampled populations. Measurements over time can be used to determine the number of rare or extinct species. Animal populations of interest can range from very large animals, such as whales (Zeh et al. 1986; Raftery & Zeh 1998), to bacteria that can only be observed under a microscope (Hong et al. 2006). Other interesting animal populations for which diversity is studied are fish (Smith & Jones 2005), fossils (Cobabe & Allmon 1994), and birds (Borgella & Gavin 2005; Walther & Martin 2001).

A problem in numismatics is to estimate the size of an ancient coin collection (Stam 1987). Esty (1986) describes that numismatists are interested in the original number of coins in an issue, and also the number of dies used to produce an ancient coinage. We can think of a die as a type of coin in the collection. The number of dies found in a sample from an ancient coin issue is used to determine the total number of dies in that coinage, which can be used to estimate the total number of coins in the issue. The size of a coin issue reveals important information regarding the economy of the civilization from which it came.

Estimating the size of an author's vocabulary is also related to estimating the

number of classes. Efron and Thisted (1976) estimate the number of words in Shakespeare's vocabulary based on word frequencies from his published material. Knowing the size of the vocabulary tells us the number of words Shakespeare knew but did not use. Also, suppose a new work by Shakespeare is discovered. An estimate of the number of new words expected to be found from this source can be used to verify authentic works. A linguist may also be interested in comparing the vocabulary size of several authors (Booth 1967).

Maintaining databases is an area of research related to estimating the number of classes. For example, when a user is searching for an journal article, it is not rare to return several hits on a search query even if the library holds only a single copy of the article. The multiple hits are all for the same published material, however multiple records are on file. These duplicate records are not identified due to discrepancies, such as a misspelling or an abbreviation in one of the search fields. The library database manager wants to know how many duplicate records are in the database, for the size of the database based on the number of records is an overestimate of the material contained in the library.

The problem of estimating the number of classes also arises in library systems management. Managers want to be informed of the usage of materials in the library and how much material is being circulated (Burrell 1988). Information about which materials are borrowed, and the frequency in which they are loaned out can be used to estimate the amount of non-circulating material.

During the debugging process of computer software, a tally of the number of errors recorded can be used to estimate the number of total faults in the system (Lloyd, Yip & Chan 1999). Errors may or may not all be equally likely to be detected. Certain resampling and debugging patterns can be used to estimate the frequency of errors in the sample. This information is used to estimate the

number of faults in the entire program.

Estimating disease incidence is also an important problem. Ding (1996) uses AIDS incidence and HIV infection data in order to estimate the size of the AIDS epidemic. Brookmeyer and Gail (1988) estimate the number of unreported diseases, as well as infection rate of AIDS. Hsieh et al. (2006) estimate the number of HIV-infected individuals from gay saunas in the Taipei area.

Böhning et al. (2004) estimates the number of illicit drug users in Bangkok. It is often difficult to gather data on any type of illegal activity, and thus difficult to make inferences on the population needed to manage many public health issues. The data collected is from hospitals that treat drug dependent patients. The number of episodes (visits) each patient makes to each facility are recorded and used to estimate the size of the drug using population.

Two early papers on estimating the number of classes were written by Good (1953) and Fisher, Corbet and Williams (1943), with the interest of estimating the frequencies of species in an animal population. Good (1953) estimates the probability of an unseen species as n_1/n where n_1 is the number of species represented by only one individual in the sample and n is the total number of observed individuals. Fisher et al. model the species abundances with a parametric gamma-mixed Poisson, or negative binomial distribution. The negative binomial model is based on assuming that the numbers of individuals from each species are independent Poisson samples, and that the means of these Poisson random variables follow a gamma distribution. Many other approaches, including Bayesian methods, have been developed for the species problem since these early works. Bayesian models for this problem are discussed in 1.3.1.

For a review on this problem including other related models and additional applications see Bunge and Fitzpatrick (1993), Buckland et al. (2000), Pollock

(2000), Schwarz and Seber (1999), and Seber and Schwarz (2002). A bibliography of related references exist and are maintained by Dr. John Bunge, Associate Professor, Cornell University. See Bunge's website, <http://www.stat.cornell.edu/~bunge/bibliography.htm>. Since the data sets used for analysis in this dissertation are biological, throughout the rest of this discussion we will refer to the classes in the population as species.

1.2 The Statistical Model

In this section, a model for the species problem is introduced which will be used to construct the likelihood function. The information for the parameters will be derived from the likelihood. In Chapters 2, 3, 4, and 5 objective priors will be derived based on the model presented in this section.

1.2.1 A Poisson Model

Let C be the unknown number of species in the population, and assume C is finite. Assume the population is infinite, and a sample is collected. Also, assume each individual, when sampled, can be uniquely identified with its respective species. Thus, for each of the C species, a nonnegative, integer number of individuals is collected in the sample. Let X_i be the abundance of species i , $i = 1, \dots, C$, or the number of individuals collected from the i th species.

The abundances are assumed to have a Poisson distribution; i.e. $X_i | \lambda_i \sim \text{Poisson}(\lambda_i)$, $i = 1, \dots, C$, where $\lambda_i > 0$ represents the mean of the Poisson distribution. We assume the X_i 's are independent. This means each species independently contributes a Poisson random number of individuals to the sample. The simplest model would consider all of the species to have equal abundances. This means for all i , $\lambda_i \equiv \lambda$, and the resulting model would be a collection of

independent and identically distributed Poisson samples.

Using an F-mixed Poisson distribution for X_i allows for considerable flexibility in the distribution of the X_i 's. We consider heterogeneity in the population by allowing the λ_i to differ between species. Specifically, differences in the abundance of each species is described by a distribution $F(\lambda|\eta)$; i.e. $\Lambda_i|\eta \sim F(\lambda|\eta)$, where $\eta = (\eta_1, \dots, \eta_m)$ represents a finite m-dimensional parameter. Therefore, the marginal distribution of X_i given η is $p_\eta(x) = \int \frac{e^{-\lambda}\lambda^x}{x!} dF(\lambda|\eta)$, representing the F-mixed Poisson distribution.

For the F-mixed Poisson distribution, the distribution $F(\lambda|\eta)$ only needs to satisfy the requirement that the support of the distribution takes positive real values, i.e. $\lambda > 0$. Taking $F(\lambda|\eta)$ as the exponential distribution, the marginal distribution of X_i given η is a geometric distribution. This model is explored in Chapter 3 with the equal abundance Poisson model as abundance distributions with one parameter. Choosing the mixing distribution to be a gamma distribution, the resulting distribution of X_i given η is a negative binomial distribution. This model is explored in Chapter 4 as an abundance distribution with two parameters. Using an abundance distribution with a larger number of parameters allows for more flexibility and the ability to model larger data sets. Our aim is to find simple abundance distributions which are versatile enough to fit a variety of data sets. In this effort, we begin to explore three parameter abundance distributions in Chapter 5 with a finite mixture of two geometric distributions. Finite mixtures of geometric distributions can be constructed as F-mixed Poisson distributions where the mixing distribution is a finite mixture of exponential distributions. In general, finite mixtures appear to be a very good modelling approach to this problem.

1.2.2 The Likelihood

In this section, the likelihood for the observed data is presented. The random abundances, (x_1, x_2, \dots, x_C) , arising from the model above are only partially observed. We will call (x_1, x_2, \dots, x_C) the *full data*. We observe the number of individuals contributed to the sample by each species, but only if the contribution is greater than zero. If a species does not appear at least once in the sample, it is not known to exist. Thus, the species that contribute zero individuals to the sample are unobserved. Since we do not observe if $x_i = 0$, the *observed data* is $n_j = \sum_{i=1}^C I(x_i = j)$ for $j \geq 1$, where $I(x_i = j)$ is 1 if $x_i = j$ and 0 otherwise. Thus, n_j represents the number of species that contribute j individuals to the sample. Species that contribute a small number of individuals are considered rare species. Species that contribute a large number of individuals are considered abundant species. The observed number of species is $w = \sum_{j \geq 1} n_j$, and the observed number of individuals is $n = \sum_{j \geq 1} j n_j$. Both w and n are random in this model and we will retain this assumption throughout. Marginally, the number of observed species, w , is a binomial random variable with size C and probability of success $1 - \Pr(X_i = 0)$. The number of individuals in the sample, n , is equal to $\sum_{i=1}^C X_i$ and its distribution depends on the distribution of the abundances. Assuming n is fixed results in a similar model to the Poisson-based model, based on the multinomial distribution. See Sandland and Cormack (1984) for a comparison of the Poisson and multinomial models. We will continue using the Poisson model throughout this dissertation. For other models related to the Poisson model presented in section 1.2.1 see Bunge and Fitzpatrick (1993).

Since we assume each Poisson sample is contributed independently, we take the likelihood to be a product of C Poisson random variables with means λ_i . By the law of total probability, we then sum over the different sets of the full data,

(x_1, x_2, \dots, x_C) , that correspond to the observed data, (n_1, n_2, \dots) . We must explicitly indicate this sum in the likelihood since the full data is unobserved. The likelihood is

$$L(\text{data}|C, \lambda_1, \dots, \lambda_C) = \sum_{x \in \mathcal{A}} \prod_{i=1}^C \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}$$

where \mathcal{A} is the set of (x_1, x_2, \dots, x_C) which correspond to the observed frequencies (n_1, n_2, \dots) .

The F-mixed Poisson model assumes heterogeneity among the mean abundances λ_i by means of a distribution dependent on a vector parameter η . Using integration to obtain the marginal distribution of the abundance counts, the integrated likelihood (Berger, Liseo & Wolpert 1999) is

$$L(\text{data}|C, \eta) = \sum_{x \in \mathcal{A}} \prod_{i=1}^C p_\eta(x_i).$$

Since $p_\eta(x_i)$ is the same for all full data vectors, we can expand the summation to obtain

$$L(\text{data}|C, \eta) = \frac{C!}{(C-w)!n_1!n_2!\dots} \prod_{i=1}^C p_\eta(x_i).$$

Sanathanan (1972) demonstrated that the likelihood can be written as

$$\begin{aligned} L(\text{data}|C, \eta) &= \binom{C}{w} (1 - p_\eta(0))^w (p_\eta(0))^{C-w} \\ &\quad \times \frac{w!}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{p_\eta(j)}{1 - p_\eta(0)} \right)^{n_j} \\ &= A(w|C, \eta) B(n_1, n_2, \dots | \eta) \end{aligned} \tag{1.1}$$

This factorization is useful since the likelihood is expressed as a function of the observed data and the likelihood is factored into two parts. The first part is a binomial likelihood for w , and the second part is a multinomial-like likelihood for the observed frequencies. The advantage to this factorization is that the

likelihood is expressed in terms of the observed data. The observed frequencies are involved in the second part of the likelihood, and $w = \sum_{j \geq 1} n_j$ is involved in the first part of the likelihood. Also, C only appears in the first part of the likelihood. The parameters in the likelihood are C and $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ and we consider η a nuisance parameter since our interest is in estimating C . Estimating C and η is considered a nonstandard problem due to the fact that the support of the data depends on C and that C is an integer valued parameter.

1.2.3 Information

The main results in this dissertation are in deriving objective priors, specifically Jeffreys and reference priors. These priors are derived from the information matrix for the parameters. The usual Fisher information is typically defined for likelihoods which are differentiable with respect to the parameters. For the species likelihood, C is a discrete parameter taking values $C = 1, 2, \dots$. This likelihood is not differentiable in C . However, Lindsay and Roeder (1987) define information for discrete parameters using the linear difference score defined as $U(N) := \frac{L(N) - L(N-1)}{L(N)}$, where $L(N)$ is the likelihood for an integer parameter N . If $U(N)$ satisfies the form $U(N) = (Y - \mu_N)/c_N$, where μ_N and c_N are functions of N and Y is random data, then the likelihood is said to have the linear difference score property.

When the linear difference score property holds, “[t]he difference score mimics in the integer parameter setting the role of the usual score function in a continuous parameter model.” (Lindsay & Roeder 1987) The linear difference score has many properties in common with the usual score function such as $E[U(N)] = 0$ and that $U(N)$ can be used to find the maximum likelihood estimate of N . When $U(N)$ has a continuous extension, the maximum likelihood estimate can be found

by solving $U(\hat{N}) = 0$.

The linear difference score for the species likelihood can be shown to satisfy the linear difference score property. Notice that $U(C)$ is the same linear difference score for the binomial model since the second, multinomial-like part of the likelihood appears in the numerator and denominator of the ratio. Thus,

$$\begin{aligned}
U(C) &= \frac{L(C, \eta) - L(C - 1, \eta)}{L(C, \eta)} \\
&= \frac{A(C, \eta)B(\eta) - A(C - 1, \eta)B(\eta)}{A(C, \eta)B(\eta)} \\
&= \frac{\binom{C}{w}(1 - p_\eta(0))^w(p_\eta(0))^{C-w} - \binom{C-1}{w}(1 - p_\eta(0))^w(p_\eta(0))^{C-1-w}}{\binom{C}{w}(1 - p_\eta(0))^w(p_\eta(0))^{C-w}} \\
&= \frac{w - C(1 - p_\eta(0))}{Cp_\eta(0)}. \tag{1.2}
\end{aligned}$$

The binomial likelihood attains the difference score property showing that the species likelihood attains this property as well.

The information about N is defined to be $1/\text{var}_N[U(N)]$. Using the methods in Lindsay and Roeder to determine the information matrix for multiparameter models with integer valued and continuous parameters, the results are the same as the asymptotic results in Sanathanan (1972). Assuming some regularity conditions (see Sanathanan 1972), using our notation, the distribution of $(C^{-1/2}(\hat{C} - C), C^{1/2}(\hat{\eta} - \eta))$ is asymptotically normal $N(0, \Sigma)$. For η having dimension m , Σ^{-1} is the $(m + 1) \times (m + 1)$ matrix

$$\Sigma^{-1} = \begin{pmatrix} \frac{1 - p_\eta(0)}{p_\eta(0)} & \left(-\frac{\partial}{\partial \eta} \log p_\eta(0)\right)^T \\ -\frac{\partial}{\partial \eta} \log p_\eta(0) & \varrho(\eta) \end{pmatrix} \tag{1.3}$$

where $\frac{\partial}{\partial \eta} \log p_\eta(0)$ is the column vector of partial derivatives,

$$\left(\frac{\partial}{\partial \eta_1} \log p_\eta(0), \frac{\partial}{\partial \eta_2} \log p_\eta(0), \dots, \frac{\partial}{\partial \eta_m} \log p_\eta(0)\right),$$

and $\varrho(\eta) = E_X \left[\left(\frac{\partial}{\partial \eta} \log p_\eta(X) \right)^2 \right]$ is $(m \times m)$ where expectation is taken with respect to p_η . The lower right partition of this matrix is the information for C random variables independent and identically distributed (i.i.d.) with distribution p_η .

We now calculate the information for the species problem using the more general method of Lindsay and Roeder. By (1.2) the score function for C is

$$U(C, \eta) = \frac{w - C(1 - p_\eta(0))}{C p_\eta(0)}.$$

The score function in η is

$$\begin{aligned} V(C, \eta) &= \frac{\partial}{\partial \eta} \log L(C, \eta) \\ &= \frac{\partial}{\partial \eta} \left[\log(C!) - \log((C - w)!) - \log\left(\prod_{j \geq 1} n_j!\right) + (C - w) \log p_\eta(0) \right. \\ &\quad \left. + \sum_{j \geq 1} \log p_\eta(j) \right] \\ &= (C - w) p_\eta^{-1}(0) p_\eta'(0) + \sum_{j \geq 1} n_j p_\eta^{-1}(j) p_\eta'(j) \end{aligned}$$

The variance of the score function in C is

$$\begin{aligned} \text{Var}(U(C, \eta)) &= \frac{C(1 - p_\eta(0)) - (C - 1)(1 - p_\eta(0))}{C p_\eta(0)} \\ &= \frac{1 - p_\eta(0)}{C p_\eta(0)} = F(C, \eta)_{11} \end{aligned} \tag{1.4}$$

For the off-diagonal elements,

$$\begin{aligned} E[U(C, \eta)V(C, \eta)] &= -E\left[\frac{\partial}{\partial \eta} U(C, \eta)\right] \\ &= -E\left[\frac{\partial}{\partial \eta} \frac{w - C(1 - p_\eta(0))}{C p_\eta(0)}\right] \\ &= -E\left[-\left(\frac{w}{C} - 1\right) p_\eta(0)^{-2} p_\eta(0)'\right] \\ &= -p_\eta(0)^{-1} p_\eta(0)' = F(C, \eta)_{12} = F(C, \eta)_{21} \end{aligned}$$

where we use $E[w] = C(1 - p_\eta(0))$ and $p_\eta(0)' = \frac{\partial}{\partial \eta} p_\eta(0)$. Finally, to complete the information matrix, we have

$$\begin{aligned}
-E \left[\frac{\partial}{\partial \eta} V(C, \eta) \right] &= -E \left[\frac{\partial}{\partial \eta} \sum_{j \geq 0} n_j p_\eta^{-1}(j) p_\eta'(j) \right] \\
&= -C \left[\sum_{j \geq 1} p_\eta(j) (p_\eta^{-1}(j) p_\eta''(j) - (p_\eta'(j))^2 p_\eta^{-2}(j)) \right] \\
&= -C E_X [p_\eta^{-1}(X) p_\eta''(X) - (p_\eta'(X))^2 p_\eta^{-2}(X)] \\
&= C \varrho(\eta) = F(C, \eta)_{22}
\end{aligned}$$

where the expectation, E_X is taken with respect to the F-mixed Poisson distribution, p_η .

The off-diagonal elements can be derived by taking the difference in C and the derivative of η (in any order). The off-diagonal elements can alternatively be written as

$$-E \left[\frac{\partial}{\partial \theta} \nabla_C \log L(C, \theta) \right] = -E \left[\nabla_C \frac{\partial}{\partial \theta} \log L(C, \theta) \right] \quad (1.5)$$

where ∇_C denotes the backward difference operation with respect to C . The left side of (1.5) is

$$\begin{aligned}
-E \left[\frac{\partial}{\partial \theta} \nabla_C \log L(C, \theta) \right] &= -E \left[\frac{\partial}{\partial \theta} (\log L(C, \theta) - \log L(C - 1, \theta)) \right] \\
&= -E \left[\frac{\partial}{\partial \theta} \left(\log \frac{C!}{(C - w)!} + (C - w) \log p_\theta(0) \right. \right. \\
&\quad \left. \left. - \log \frac{(C - 1)!}{(C - 1 - w)!} - (C - 1 - w) \log p_\theta(0) \right) \right] \\
&= -E \left[\frac{\partial}{\partial \theta} \log p_\theta(0) \right] \\
&= -\frac{\partial}{\partial \theta} \log p_\theta(0).
\end{aligned}$$

The right side of (1.5) is

$$\begin{aligned}
-E \left[\nabla_C \frac{\partial}{\partial \theta} \log L(C, \theta) \right] &= -E \left[\nabla_C \left((C - w) \frac{\partial}{\partial \theta} \log p_\theta(0) \right. \right. \\
&\quad \left. \left. + \sum_{j \geq 1} n_j \frac{\partial}{\partial \theta} \log p_\theta(j) \right) \right] \\
&= -E \left[(C - w) \frac{\partial}{\partial \theta} \log p_\theta(0) \right. \\
&\quad \left. - (C - 1 - w) \frac{\partial}{\partial \theta} \log p_\theta(0) \right] \\
&= -E \left[\frac{\partial}{\partial \theta} \log p_\theta(0) \right] \\
&= -\frac{\partial}{\partial \theta} \log p_\theta(0).
\end{aligned}$$

This holds for a vector, $\theta = (\theta_1, \theta_2, \dots, \theta_m)$. In that case consider

$$\frac{\partial}{\partial \theta} \log L(C, \theta) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} \log L(C, \theta) \\ \frac{\partial}{\partial \theta_2} \log L(C, \theta) \\ \vdots \\ \frac{\partial}{\partial \theta_m} \log L(C, \theta) \end{pmatrix}$$

Then, $-E \left[\nabla_C \frac{\partial}{\partial \theta} \log L(C, \theta) \right]$ is $(m \times 1)$.

Therefore, the information for C and η is

$$F(C, \eta) = \begin{pmatrix} \frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} & \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T \\ -\frac{\partial}{\partial \eta} \log p_\eta(0) & C \varrho(\eta) \end{pmatrix} \quad (1.6)$$

Notice that the diagonal elements of this partitioned matrix contain elements which factor into a function of C times a function of η , the nuisance parameter. A factorization in the elements of the information matrix will become important in later chapters in deriving the prior for C and η . Also, note that the information is written as a function of the abundance distribution p_η . This also simplifies calculations since our derivations will involve p_η instead of the entire likelihood function.

1.3 An Objective Bayesian Approach

This section will describe and motivate the Bayesian approach taken in this dissertation. We will begin by reviewing Bayesian methods for the species problem, and then discuss why using objective priors is useful, appropriate, and advantageous.

Among all objective priors, we wish to find a prior that is automatically selected for our problem. The rules used to derive the priors in this dissertation are model-based; that is, they utilize information from the model.

1.3.1 Bayesian Estimation for the Species Problem

Bayesian estimation for the species problem involves placing a prior on the number of species and the parameters of the abundance distribution. We will denote the prior for the number of species and the prior for the nuisance parameter as $\pi(C)$ and $\pi(\eta)$, respectively. The joint prior will be denoted $\pi(C, \eta)$.

Hill (1979) uses a negative binomial prior for C and estimates the posterior probability for sampling a previously unobserved species. Lewins and Joanes (1984) also use a negative binomial distribution for the marginal prior of C . For the two parameters in the negative binomial distribution, the authors suggest expert advice in specifying these values. A symmetric Dirichlet prior is used for the relative abundances. For the Dirichlet parameters, the authors suggest specifying the parameter carefully, or placing a hyperprior on this parameter. Caution is advised to be taken since different values for the Dirichlet parameter change the posterior substantially. A way to estimate the Dirichlet parameter from the repeat rate in the data is also shown. However, this empirical method to estimate the parameter in the prior distribution is dependent on the observed number of species.

Blumenthal (1992) discusses how Bayesian estimates have been investigated as a solution to the instability of estimators for C . In Blumenthal and Marcus (1975), applications to truncated data are focused on the continuous, exponential distribution as opposed to the discrete, Poisson model described in section 1.2.1; and use of a proper uniform distribution for C is evaluated in terms of expected bias. Blumenthal also compares properties of the unconditional MLE, the conditional MLE, and variations on the MLE based on asymptotic differences, such as bias, MSE, and second and higher order terms (Blumenthal 1977; Blumenthal, Dahiya, & Gross 1978).

Boender and Rinooy Kan (1987) use a model based on the generalized multinomial distribution in order to estimate the number of species and the coverage of the sample. An arbitrary prior is used for C to give results. A symmetric Dirichlet distribution is used for the multinomial cell probabilities and a prior for the parameter of the Dirichlet is also used.

Solow (1994) reviews a group of models that follow a Bayesian approach. The symmetric Dirichlet and the sequential broken stick model are used to model the prior probability for the relative abundances. In the symmetric Dirichlet prior, the parameter known as the ‘flattening constant’ can be estimated from the data in an empirical Bayes approach, or the parameter can be specified with a prior distribution. Solow argues that the sequential broken stick model has ‘considerable empirical support’. The sequential broken stick model’s distribution has no closed form and no parameters. The use of sequential broken stick model and symmetric Dirichlet are compared using two data sources. The prior used for C is a negative binomial distribution with fixed parameter values.

Rodrigues, Milan and Leite (2001) compare a fully Bayes method with an empirical Bayes method for estimating the number of species. The prior $\pi(C) \propto 1/C$

is specified for the total number of species. The data is assumed to be Poisson. A gamma prior is placed on the Poisson means. The hyperparameters are given objective improper priors. Full conditionals are derived for the number of unseen species and the two parameters of the abundance distribution. Empirical Bayes estimates based on conditional and profile likelihoods are shown. “[The] empirical Bayes approach approximates to the fully Bayesian approach for large values of C .” The authors also summarize advantages of an empirical Bayes approach including its appeal to non-Bayesians. In a numerical example a poor agreement between marginal posteriors for C based on empirical and fully Bayesian methods is shown. This difference is attributed to small sample sizes.

Wang, He and Sun (2007) compare four objective priors with a capture model with capture probabilities varying among sampling occasions. The improper priors $\pi(C) \propto 1$ and $\pi(C) \propto 1/C$ are compared with maximum likelihood estimates and are found to perform better for small sample sizes. They also discuss the requirements for the posterior distribution to be proper when using these improper priors on C , which reduce to conditions on the sample size.

Rissanen’s prior (1983) is a proper objective prior based on coding theory and is used by Tardella (2002) and Madigan and York (1997). Using an optimal, prefix code for the integers, we can define the probability of each integer by its codeword length. The Kraft inequality is used to prove this prior is proper. The prior takes the form $\pi(C) = 2^{-Q(C)}$, where $Q(C) = \log C + \log \log C + \dots$ and only the positive terms of $Q(C)$ are summed. Unfortunately, the technique used to derive this prior is not extendable to multiparameter models.

Among the previous sources (and some others) there have been several suggestions and implementations of priors on C . Priors that have been proposed for C are the negative binomial (Hill 1979; Rodrigues 2001; Raftery 1987; Wang

et al. 2007; Lewins & Joanes 1984), Poisson (Madigan 1997; Raftery 1988), improper uniform (Tardella 2002; Boender & Rinnooy Kan 1987; Wang et al. 2007), and Jeffreys' proposed prior, $\pi(C) \propto 1/C$, known as Jeffreys prior (Jeffreys 1939/1961). Jeffreys's proposed prior also results from hierarchical structure where $\pi(C|\lambda)$ is Poisson (λ) and $\pi(\lambda) \propto \lambda^{-1}$ (Raftery 1988). Jeffreys' proposed prior has been widely used (Wang et al. 2007; George & Robert 1992; Smith 1991; Tardella 2002; Madigan & York 1997). Priors of the form $\pi(C) \propto 1/C^r$, where r is a nonnegative constant have been described in Wang et al. (2007). This class of priors for C includes the improper uniform prior, when $r = 0$ and Jeffreys' proposed prior when $r = 1$.

The Bayesian approach for this problem has several advantages. As opposed to maximum likelihood estimates, the Bayesian method naturally produces a posterior distribution which is discrete, lending itself easily to integer valued point and interval estimates. Also, posterior interval estimates are restricted to be within the parameter space.

All of the priors above assume C and η are independent. A fundamental difference in the priors proposed in this dissertation and all of the above methods, is that the priors are derived jointly instead of being assumed a priori independent. The Jeffreys and reference rules allow for this derivation. Nonetheless, due to the form of the likelihood, the joint prior factors into a product of two independent priors for C and η .

1.3.2 Objective Priors

Within a Bayesian framework, we choose to use objective, or noninformative priors. For an overview on noninformative priors see Kass and Wasserman (1996). In many cases prior information is unavailable or controversial. Objective priors

can also be used in a sensitivity analysis for comparison with other subjective priors. In this dissertation we present a prior for the model parameters which is jointly derived based on a single notion of objectiveness. Methods of deriving objective priors by Bernardo (1979) and Jeffreys (1946) are used to jointly derive a multivariate prior. This procedure does not assume the priors are a priori independent. One disadvantage of objective priors is that they are often improper, and so we are required to show that the posterior integrates to unity in order to allow valid inference.

Jeffreys Prior

The Jeffreys prior (Jeffreys 1946) is considered an objective prior and is based on invariance under one-to-one reparameterization. The principle is that any prior for a parameter ϕ should yield an equivalent result if a model with a transformed parameter is used. The Jeffreys prior is defined to be proportional to the square root of the Fisher information. This means for a scalar parameter ϕ and model $p(\mathbf{x}|\phi)$, the Jeffreys prior is

$$\pi(\phi) \propto h(\phi)^{1/2} \tag{1.7}$$

where

$$h(\phi) = \int_X p(\mathbf{x}|\phi) \left(\frac{\partial}{\partial \phi} \log p(\mathbf{x}|\phi) \right)^2 dx. \tag{1.8}$$

For multidimensional models, the determinant of the Fisher information matrix can be used, which preserves the invariance property. The Jeffreys prior is often improper, so integrability of the posterior must be shown when using this prior. The one-dimensional Jeffreys prior can be derived from several different approaches (Bernardo & Smith 2000) and has been widely used.

The use of Jeffreys prior in multivariate models is controversial. Use of independent priors, where Jeffreys' rule is applied to each parameter separately,

has been recommended. Other objective priors are preferred and often yield preferable results.

Reference Prior

Bernardo's reference prior (Bernardo 1979; Bernardo & Ramón 1998) is an objective prior which is based on maximizing the expected entropy provided by the prior. This can also be interpreted as maximizing the missing information from the experiment. Information is defined in the same way as Shannon entropy (Lindley 1956). For a prior distribution $p(\phi)$, the amount of information is defined to be

$$\mathfrak{g}_0 = \int p(\phi) \log p(\phi) d\phi.$$

For a posterior distribution $p(\phi|\mathbf{x})$, the amount of information after \mathbf{x} has been observed is

$$\mathfrak{g}_1 = \int p(\phi|\mathbf{x}) \log p(\phi|\mathbf{x}) d\phi$$

The amount of information provided by the experiment ε , with prior knowledge $p(\phi)$ is

$$\mathfrak{g}(\varepsilon, p(\phi), \mathbf{x}) = \mathfrak{g}_1 - \mathfrak{g}_0$$

The average amount of information provided by the experiment ε , with prior knowledge $p(\phi)$, is

$$\begin{aligned} I\{\varepsilon, p(\phi)\} &= E_{\mathbf{x}} [\mathfrak{g}_1 - \mathfrak{g}_0] \\ &= \int_X p(\mathbf{x}) \int_{\Phi} p(\phi|\mathbf{x}) \log \left(\frac{p(\phi|\mathbf{x})}{p(\phi)} \right) d\phi dx \end{aligned} \quad (1.9)$$

The reference prior is derived by minimizing (1.9) with respect to $p(\phi)$. The derivation of the reference prior takes into account the order of interest of the

parameters, namely for our problem, that C is the parameter of interest and η is a nuisance parameter. In the case of multiple nuisance parameters an ordering must be assigned; however, in specific cases it has been shown that this ordering is not important regardless that there is no guarantee to obtain a unique reference prior from different orderings (Irony 1997).

For a model $p(\mathbf{x}|\phi, \lambda)$ with interest parameter ϕ and nuisance parameter λ , the joint prior $\pi(\phi, \lambda)$ is derived in steps. First, a conditional prior $\pi(\lambda|\phi)$ is derived from the model. Then, the marginal model $p(\mathbf{x}|\phi)$ is obtained after integrating out the nuisance parameter, and this model is used to find the prior $\pi(\phi)$.

The reference prior satisfies many desirable properties for objective priors (Bernardo & Ramón 1998). Like the Jeffreys prior, the reference prior is invariant to reparameterization. It turns out for one dimensional problems, when there is only one unknown parameter, the reference prior is equivalent to the Jeffreys prior.

1.4 Data Sets

Two data sets will be used as examples to implement the Bayesian procedure described in this dissertation. The results will be used to evaluate the method as well as compare to other analyses.

1.4.1 Framvaren Fjord

The first data set is a microbial sample from the Framvaren Fjord in Norway. The Framvaren Fjord is described as an anoxic (oxygen-depleted) environment. This extreme environment is of interest to researchers in understanding diversity of microorganisms. Diversity of these organisms is largely unknown See Behnke et al. (2006) for details. Water samples were taken 18 meters below the water's

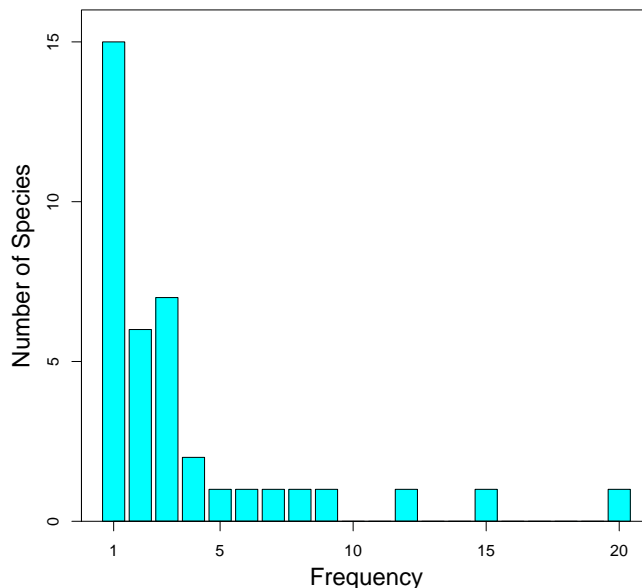


Figure 1.1: Plot of frequency counts for the Framvaren Fjord data. Frequencies 1 through 20 are shown.

surface. The organisms from the sample were classified into species based on their 18S rRNA similarity.

The observed frequencies are shown in Table 1.1 and a plot of the frequency counts (for observed frequencies less than or equal to twenty) is in Figure 1.1. For the full data, the observed number of species is $w = 39$ and the observed number of individual organisms is $n = 303$.

1.4.2 Lepidoptera

The second data set we use is from a capture experiment on Lepidoptera (Fisher et al. 1943). A sample of Lepidoptera was collected in Harpenden, England. The insects were collected in traps over four years (1933-1936). The number of insects of each species were counted. There are $w = 240$ observed species and $n = 15,609$ observed individual insects. Table 1.2 shows the abundance counts for some of

Table 1.1: Frequency counts for Framvaren Fjord

Frequency	Count
1	15
2	6
3	7
4	2
5	1
6	1
7	1
8	1
9	1
12	1
15	1
20	1
165	1

the species from this experiment (Williams 1939). Table 1.3 and Figure 1.2 show the frequency counts.

1.5 Layout of the Dissertation

Chapter 2 discusses the case when the nuisance parameter η is known. The likelihood reduces to a binomial likelihood. Although this situation does not

Table 1.2: Abundance counts for species of Lepidoptera

Species	Abundance
<i>Selenia lunaria</i>	15
<i>Phegalia pedaria</i>	2
<i>Lygris associata</i>	4
<i>Agrochola circellaris</i>	2
<i>Agrotis exclamationis</i>	2349
<i>Agrotis strigula</i>	1
\vdots	\vdots
<i>Melanchra pisi</i>	4

Table 1.3: Frequency counts for species of Lepidoptera

Frequency	Number of Species
1	35
2	11
3	15
4	14
5	10
6	11
7	5
8	6
\vdots	\vdots
2349	1

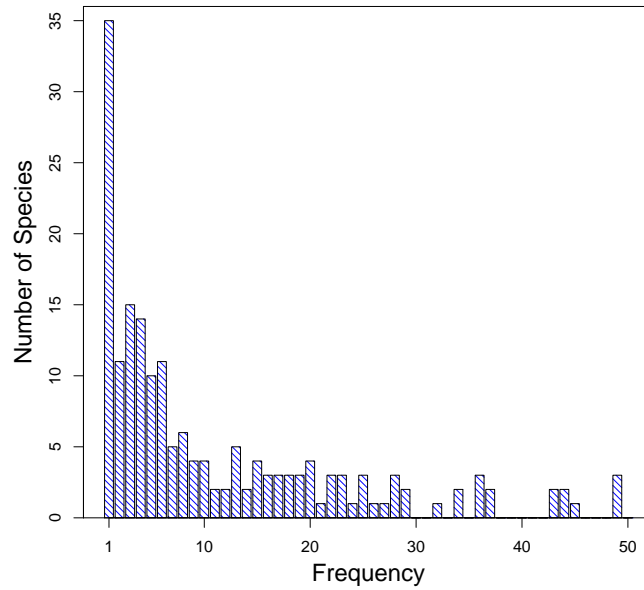


Figure 1.2: Plot of frequency counts for Lepidoptera data. Frequencies 1 through 50 are shown.

have practical use for our problem, the scenario represents the model having zero nuisance parameters. Chapter 3 begins with the simplest practical models, the equal abundance Poisson and geometric models. These distributions represent models with one nuisance parameter. We derive the joint Jeffreys and reference priors for these models and implement them using two data sets. Chapter 4 moves to models with two nuisance parameters, which includes the negative binomial distribution. The joint Jeffreys and reference priors are derived. Implementation on the example data sets is shown using an alternative to the joint Jeffreys and reference priors which guarantee proper posteriors. Chapter 5 discusses models with three or more nuisance parameters. Forms for the Jeffreys and reference priors are shown for a general finite dimensional η . Implementation is shown for a model defined by a finite mixture of two geometric distributions. Chapter 6 provides a comparison of the Jeffreys and reference priors for this problem as well as discussion on other frequentist and nonparametric methods. A procedure for a fully Bayesian approach to estimating the number of classes in a population is discussed.

Chapter 2

Estimation Conditional on the Nuisance Parameters

This chapter considers the likelihood for the species problem with the simplifying assumption that the nuisance parameter η is known. This special case is useful in illustrating our techniques with a simplified model and relates to the more complicated models in later chapters.

2.1 Jeffreys and Reference Priors

The second part of the likelihood, $B(n_1, n_2, \dots | \eta)$, in (1.1) does not depend on C . Therefore, assuming η is known reduces the likelihood to a binomial distribution. We will use an alternate notation in this chapter to distinguish this special case from the models in later chapters. Denote the likelihood as

$$L(x|N) = \binom{N}{x} p^x (1-p)^{N-x}$$

where x is data and p is known, Lindsay and Roeder (1987) show the information for N is $F(N) = p/\{N(1-p)\}$. Replacing the Fisher information with the information in N yields the Jeffreys prior as $\pi(N) \propto (1/N)^{1/2}$. This is an improper prior; i.e., $\sum_{N=1}^{\infty} (1/N)^{1/2} = \infty$. Since the Jeffreys and reference prior are the same for problems with one unknown parameter (Bernardo 2000) the Jeffreys

prior and reference prior are both $\pi(N) \propto (1/N)^{1/2}$.

In order to use this prior we need to show it yields a proper posterior. We can prove finite integrability since

$$\begin{aligned}\pi(N|x) &\propto \pi(N)L(x|N) \\ &= \frac{1}{\sqrt{N}} \binom{N}{x} p^x (1-p)^{N-x} \\ &\propto \sqrt{N} \binom{N-1}{x-1} p^x (1-p)^{N-x}.\end{aligned}$$

The posterior is proper since $\pi(N|x) \propto E[N^{1/2}]$ where expectation is taken with respect to a negative binomial distribution. Since the negative binomial distribution has all finite moments, the posterior is proper.

Since Jeffreys prior is based on the invariance principle, let us consider a one-to-one transformation of N . Let $\gamma = h(N)$ be a one-to-one reparameterization of N in the binomial model. Then

$$L(x|\gamma) = \binom{h^{-1}(\gamma)}{x} p^x (1-p)^{h^{-1}(\gamma)}.$$

Consider the linear difference score for γ . We have

$$U(\gamma) = \frac{L(\gamma) - L(\gamma - 1)}{L(\gamma)}.$$

After a reparameterization of the model, there is no guarantee that the likelihood has the linear difference score property. For example, consider a transformation that depends on x . Specifically, suppose $h(N) = N - x$. Then,

$$U(\gamma) = \frac{x - (\gamma + x)p}{(\gamma + x)(1-p)}$$

and the likelihood no longer has the linear difference score property. This example shows that it is not trivial to consider transformations of the model.

Showing invariance of the Jeffreys prior in general for discrete parameter models using the linear difference score is a problem that poses several mathematical

difficulties. Transformations of the score function may not satisfy the linear difference score property. It is also possible to consider N as a continuous parameter in the model, however, complications arise in showing regularity conditions hold for the Fisher information. Also, it is not clear that transformations for our model are practical. For the species problem we will only consider the standard parameterization of the binomial distribution.

2.2 A Class of Priors for the Binomial Likelihood

The prior for N derived in section 2.1 is within a class of priors taking the form $\pi(N) \propto 1/N^\alpha$ with $\alpha \in [0, 1]$, which all yield proper posteriors. Let

$$\pi(N) = \frac{1}{N^\alpha} I(N \in \{1, 2, \dots\})$$

where $I(N \in \mathcal{S})$ is 1 if N is in the set \mathcal{S} and 0 otherwise. The posterior is

$$\begin{aligned} \pi(N|x) &\propto \pi(N)L(x|N) \\ &\propto \frac{1}{N^\alpha} \binom{N}{x} p^x (1-p)^{(N-x)} I(N \in \{x, x+1, \dots\}) \\ &= \frac{N^{(1-\alpha)}}{x} \binom{N-1}{x-1} p^x (1-p)^{(N-x)} I(N \in \{x, x+1, \dots\}) \\ &\propto N^{(1-\alpha)} \binom{N-1}{x-1} p^x (1-p)^{(N-x)} I(N \in \{x, x+1, \dots\}) \end{aligned} \quad (2.1)$$

We can show that the posterior is proper since

$$\begin{aligned} \sum_{N=x}^{\infty} \pi(N|x) &= K \sum_{N=x}^{\infty} N^{(1-\alpha)} \binom{N-1}{x-1} p^x (1-p)^{(N-x)} \\ &\propto E[N^{(1-\alpha)}] \end{aligned} \quad (2.2)$$

where expectation in (2.2) is taken with respect to a negative binomial distribution with parameters x and p . For $\alpha = 0$, $E[N^{(1-\alpha)}] = x/p$, the first moment of

the negative binomial distribution. For any $\alpha \in (0, 1]$, each term in the summation is positive and less than its corresponding term in $E[N]$. By the comparison test, the series converges. Thus, the posterior is proper for all $\alpha \in [0, 1]$.

The mean of the posterior is

$$\begin{aligned} E[N|x] &= \sum_{N=x}^{\infty} N \pi(N|x) \\ &\propto \sum_{N=x}^{\infty} N N^{(1-\alpha)} \binom{N-1}{x-1} p^x (1-p)^{(N-x)} \\ &= E[N^{(2-\alpha)}], \end{aligned}$$

where expectation is taken with respect to a negative binomial distribution. Substituting $\alpha = 0$ in $E[N^{(2-\alpha)}]$ corresponds to the second moment of a negative binomial distribution, which is finite. For any $\alpha \in (0, 1]$, each term in $E[N^{(2-\alpha)}]$ is less than the corresponding term in $E[N^2]$. Again, by the comparison test, the series converges. Thus, the posterior has a finite mean.

The posterior has a finite variance. The second moment of the posterior is

$$\begin{aligned} E[N^2|x] &= \sum_{N=x}^{\infty} N^2 \pi(N|x) \\ &\propto E[N^{(3-\alpha)}] \end{aligned}$$

which, by a similar argument as above, is finite. In fact, for $k \geq 1$, the k -th moment is finite, since we can write the k -th moment as proportional to $E[N^{(k+1)-\alpha}]$, which is the $(k+1)-\alpha$ moment of the negative binomial distribution.

Also note that the posterior for N and its moments are finite for $\alpha > 1$.

Special case: $\alpha = 1/2$

The noninformative prior for $\alpha = 1/2$ is

$$\pi(N) = \frac{1}{N^{1/2}} I(N \in \{1, 2, \dots\})$$

The posterior mode can be found using

$$\begin{aligned}\frac{f(N+1|x)}{f(N|x)} &= \frac{(N+1)^{1/2} \binom{N}{x-1} p^x (1-p)^{(N+1-x)}}{N^{1/2} \binom{N-1}{x-1} p^x (1-p)^{(N-x)}} \\ &= \left(\frac{N+1}{N} \right)^{1/2} \frac{N}{N-x+1} (1-p).\end{aligned}$$

Setting $\frac{f(N+1|x)}{f(N|x)}$ equal to 1 gives

$$[N(N+1)]^{1/2} (1-p) = N-x+1$$

Solving the above equation for N and rounding up to the nearest integer will give us the mode. This gives a unique solution when $p \in (0, 1)$ and $x > 0$.

Special case: $\alpha = 0$

The prior for $\alpha = 0$ is

$$\pi(N) = I(N \in \{1, 2, \dots\})$$

This is a constant function on the positive integers. The posterior is

$$\pi(N|x) = \binom{N}{x} p^{(x+1)} (1-p)^{(N-x)} I(N \in \{x, x+1, \dots\})$$

with mean,

$$\begin{aligned}E[N|x] &= \sum_{N=x}^{\infty} N \binom{N}{x} p^{(x+1)} (1-p)^{(N-x)} \\ &= \frac{x+1-p}{p} = \frac{x}{p} + \frac{1-p}{p}\end{aligned}$$

and variance

$$\begin{aligned}Var(N|x) &= E[N^2|x] - (E[N|x])^2 \\ &= \frac{p^2 - 3xp - 3p + x^2 + 2 + 3x}{p^2} - \left(\frac{x+1-p}{p} \right)^2 \\ &= \frac{x - px - p + 1}{p^2} = \frac{x}{p} \frac{1-p}{p} + \frac{1-p}{p^2}.\end{aligned}$$

Also the mode is the smallest integer N such that $\pi(N+1|x) > \pi(N|x)$, or the smallest integer N not less than the solution to $\pi(N+1|x)/\pi(N|x) = 1$, and

$$\begin{aligned}\frac{\pi(N+1|x)}{\pi(N|x)} &= \frac{\binom{N+1}{x} p^{(x+1)} (1-p)^{(N+1-x)}}{\binom{N}{x} p^{(x+1)} (1-p)^{(N-x)}} \\ &= \frac{N+1}{N-x+1} (1-p).\end{aligned}$$

Then, setting

$$\frac{\pi(N+1|x)}{\pi(N|x)} = 1$$

implies

$$N = \frac{x-p}{p} = \frac{x}{p} - 1.$$

Thus, the posterior mode is $\lceil \frac{x}{p} - 1 \rceil$ or $\lfloor \frac{x}{p} \rfloor$.

Special case: $\alpha = 1$

The prior for $\alpha = 1$ is

$$\pi(N) = \frac{1}{N} I(N \in \{1, 2, \dots\}).$$

The posterior is

$$\begin{aligned}\pi(N|x) &\propto \frac{1}{N} \binom{N}{x} p^x (1-p)^{(N-x)} I(N \in \{x, x+1, \dots\}) \\ &\propto \binom{N-1}{x-1} p^x (1-p)^{(N-x)} I(N \in \{x, x+1, \dots\})\end{aligned}$$

which is a negative binomial distribution with parameters x and p . The posterior mean is x/p , the posterior variance is $\frac{x(1-p)}{p^2}$, and the posterior mode is $\lceil \frac{x-1}{p} \rceil$ or $\lfloor \frac{x-1-p}{p} \rfloor$. The posterior mean for this case is the same as the unbiased estimate for N obtained in a frequentist setting. This prior is often referred to as the Jeffreys prior since its appearance in Jeffreys (1939/1961). In this paper we will refer to the prior $\pi(N) \propto 1/N$ as the proposed Jeffreys prior.

Chapter 3

One Nuisance Parameter

This chapter discusses models having η as a one-dimensional nuisance parameter. The equal abundance Poisson model and the geometric model both have one-dimensional nuisance parameters. These two specific models will be discussed in this chapter. First, let us recall the model we are using where p_η is a one-dimensional abundance distribution.

The likelihood rewritten from (1.1) is

$$\begin{aligned} L(\text{data}|C, \eta) &= \binom{C}{w} (1 - p_\eta(0))^w (p_\eta(0))^{C-w} \\ &\quad \times \frac{w!}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{p_\eta(j)}{1 - p_\eta(0)} \right)^{n_j} \\ &= \frac{C!}{(C-w)!} (p_\eta(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \end{aligned} \quad (3.1)$$

where (3.1) shows an equivalent form which is often simpler to use in calculations.

The information matrix from (1.6) for the parameters (C, η) is equal to

$$F(C, \eta) = \begin{pmatrix} \frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} & -\frac{\partial}{\partial \eta} \log p_\eta(0) \\ -\frac{\partial}{\partial \eta} \log p_\eta(0) & C \varrho(\eta) \end{pmatrix} \quad (3.2)$$

where $\varrho(\eta) = E_X \left[\left(\frac{\partial}{\partial \eta} \log p_\eta(X) \right)^2 \right]$.

3.1 Jeffreys Prior

In this section the prior for (C, η) will be derived using the multivariate Jeffreys' rule described in section 1.3.2 replacing the Fisher information with the information in (3.2).

The determinant of the information matrix in (3.2) is

$$\begin{aligned}\det(F(C, \eta)) &= \left(\frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} \right) \left(C \varrho(\eta) \right) - \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^2 \\ &= \frac{1 - p_\eta(0)}{p_\eta(0)} \varrho(\eta) - \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^2\end{aligned}$$

Thus, the Jeffreys prior is

$$\pi(C, \eta) \propto \left[\frac{1 - p_\eta(0)}{p_\eta(0)} \varrho(\eta) - \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^2 \right]^{1/2} \quad (3.3)$$

The Jeffreys prior for C and η factors into two independent priors. The marginal prior for C , $\pi(C)$ is proportional to a constant for $C = 1, 2, \dots$

Since $\pi(C, \eta)$ is an improper prior (although $\pi(\eta)$ may be integrable over values of η), we want to show the posterior is proper. This is difficult to prove in full generality. However, later in this chapter the posterior resulting from the Jeffreys prior is shown to be proper in two specific cases.

The posterior for the model using the prior $\pi(C, \eta)$ is

$$\begin{aligned}\pi(C, \eta | \text{data}) &\propto \pi(C, \eta) L(\text{data} | C, \eta) \\ &= \pi(\eta) \frac{C!}{(C - w)!} (p_\eta(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j}\end{aligned}$$

where $C = w, w + 1, w + 2, \dots$ and $\eta \in \mathbb{R}$. We need to show

$$\int d\pi(C, \eta | \text{data}) < \infty.$$

We begin with the iterated integral

$$\begin{aligned}
& \int_{\mathbb{R}} \sum_{C \geq w} \pi(C, \eta | \text{data}) d\eta \\
&= \int_{\mathbb{R}} \sum_{C \geq w} \pi(\eta) \frac{C!}{(C-w)!} (p_\eta(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \\
&= \int_{\mathbb{R}} \pi(\eta) \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \sum_{C \geq w} \frac{C!}{(C-w)!} (p_\eta(0))^{C-w}
\end{aligned}$$

and the part of the posterior which is dependent on C is a negative binomial distribution. Thus, we can sum over C obtaining

$$\sum_{C \geq w} \frac{C!}{(C-w)!} (p_\eta(0))^{C-w} = \frac{w!}{(1 - p_\eta(0))^{w+1}}$$

and the marginal posterior for η is

$$\pi(\eta | \text{data}) \propto \pi(\eta) \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \frac{1}{(1 - p_\eta(0))^{w+1}}.$$

When considering a model after specifying p_η , integrability of the posterior can be shown if $\int \pi(\eta | \text{data}) d\eta < \infty$.

Using the derivations above, we will look closely at the Poisson and geometric models. The form of the Jeffreys priors will be examined and the priors will be shown to yield proper posterior distributions.

3.1.1 Poisson Model

Suppose $p_\eta(x)$ is a Poisson distribution with mean $\lambda > 0$; thus

$$p_\lambda(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

where $x = 0, 1, 2, \dots$. This model assumes that all species are equally abundant in the population. The likelihood from (3.1) written for the Poisson model is

$$\begin{aligned}
L(\text{data} | C, \lambda) &= \frac{C!}{(C-w)!} (e^{-\lambda})^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{e^{-\lambda} \lambda^j}{j!} \right)^{n_j} \\
&= \frac{C!}{(C-w)!} (e^{-\lambda})^C \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{1}{j!^{n_j}} \right) \lambda^n.
\end{aligned}$$

Using (3.2), the (2, 2) element of the information matrix is

$$F(C, \lambda)_{22} = -C \left\{ E_X \left[\frac{\partial^2}{\partial \lambda^2} \log p_\lambda(X) \right] \right\}$$

where we use the simplifying assumption

$$E_X \left[\left(\frac{\partial}{\partial \eta} \log p_\eta(X) \right)^2 \right] = - \left\{ E_X \left[\frac{\partial^2}{\partial \eta^2} \log p_\eta(X) \right] \right\}.$$

Therefore,

$$\begin{aligned} F(C, \lambda)_{22} &= -C \left\{ E_X \left[\frac{\partial^2}{\partial \lambda^2} - \lambda + X \log \lambda - \log X! \right] \right\} \\ &= -C \left\{ E_X \left[-\frac{X}{\lambda^2} \right] \right\} \\ &= \frac{C}{\lambda} \end{aligned}$$

where the last equality is true since $E[X] = \lambda$. The off-diagonal elements of the information matrix are

$$\begin{aligned} F(C, \lambda)_{21} = F(C, \lambda)_{12} &= -\frac{\partial}{\partial \lambda} \log p_\lambda(0) \\ &= -\frac{\partial}{\partial \lambda} (-\lambda) \\ &= 1. \end{aligned}$$

The (1, 1) element of the information is

$$\begin{aligned} F(C, \lambda)_{11} &= \frac{1}{C} \frac{1 - p_\lambda(0)}{p_\lambda(0)} \\ &= \frac{1}{C} \frac{1 - e^{-\lambda}}{e^{-\lambda}} \\ &= \frac{e^\lambda - 1}{C}. \end{aligned}$$

Thus, the information for C and λ is

$$F(C, \lambda) = \begin{pmatrix} \frac{e^\lambda - 1}{C} & 1 \\ 1 & \frac{C}{\lambda} \end{pmatrix} \quad (3.4)$$

and the Jeffreys prior is

$$\pi(C, \lambda) \propto \left(\frac{e^\lambda - \lambda - 1}{\lambda} \right)^{1/2}.$$

This prior is improper, increasing in λ , and constant in C .

Using the Jeffreys prior the posterior is

$$\begin{aligned} \pi(C, \lambda | \text{data}) &\propto \pi(C, \lambda) L(\text{data} | C, \lambda) \\ &= \left(\frac{e^\lambda - \lambda - 1}{\lambda} \right)^{1/2} \frac{C!}{(C-w)!} (e^{-\lambda})^C \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{1}{j!^{n_j}} \right) \lambda^n \\ &\propto \binom{C}{w} (1 - e^{-\lambda})^{w+1} (e^{-\lambda})^{C-w} (e^\lambda - \lambda - 1)^{1/2} (1 - e^{-\lambda})^{-w-1} \\ &\quad \times e^{-\lambda w} \lambda^{n-1/2} \\ &= \text{negative binomial}(w+1, 1 - e^{-\lambda}) (e^\lambda - \lambda - 1)^{1/2} \\ &\quad \times (1 - e^{-\lambda})^{-w-1} e^{-\lambda w} \lambda^{n-1/2} \end{aligned}$$

where the negative binomial can be parameterized as

$$\binom{(C-w) + (w+1) - 1}{(w+1) - 1} (1 - e^{-\lambda})^{w+1} (e^{-\lambda})^{C-w} = \binom{C}{w} (1 - e^{-\lambda})^{w+1} (e^{-\lambda})^{C-w}$$

for $C - w = 0, 1, \dots$. This means the number of species that did not appear in the sample has a negative binomial posterior distribution with size = $w + 1$ and probability = $1 - e^{-\lambda}$.

Next, we show the posterior is proper.

Proposition 1. *The posterior for the Poisson model with Jeffreys prior given by*

$$\pi(C, \lambda | \text{data}) \propto \left(\frac{e^\lambda - \lambda - 1}{\lambda} \right)^{1/2} \frac{C!}{(C-w)!} (e^{-\lambda})^C \lambda^n$$

is proper.

Proof. Integrating over C and λ we have

$$\begin{aligned} \int d\pi(C, \lambda | \text{data}) &\propto \int_0^\infty \frac{(e^\lambda - \lambda - 1)^{1/2}}{(1 - e^{-\lambda})^w} e^{-\lambda w} \lambda^{n-1/2} \sum_{C=w}^\infty \binom{C}{w} (e^{-\lambda})^{C-w} \\ &\quad \times (1 - e^{-\lambda})^w d\lambda \\ &\propto \int_0^\infty (e^\lambda - \lambda - 1)^{1/2} e^{-\lambda w} \lambda^{n-1/2} (1 - e^{-\lambda})^{-w-1} d\lambda. \end{aligned}$$

Now, $(e^\lambda - \lambda - 1)^{1/2}$ can be bounded from above by $(e^\lambda)^{1/2}$ and $(1 - e^{-\lambda})^{-w-1}$ can be bounded from above by $B > 1$ for all $\lambda \geq \lambda_B$. Also, $(1 - e^{-\lambda})^{-w-1}$ can be bounded from above by $(2/\lambda)^{(w+1)}$ for $0 < \lambda \leq \lambda_B$. Thus, consider the integral which bounds the posterior

$$\begin{aligned} K \int_0^{\lambda_B} e^{-\lambda(w-1/2)} \lambda^{n-w-3/2} d\lambda + K^* \int_{\lambda_B}^\infty e^{-\lambda(w-1/2)} \lambda^{n-1/2} d\lambda \\ < K \int_0^\infty e^{-\lambda(w-1/2)} \lambda^{n-w-3/2} d\lambda + K^* \int_0^\infty e^{-\lambda(w-1/2)} \lambda^{n-1/2} d\lambda \\ < \infty \end{aligned}$$

as long as $w \geq 1$ and $n \geq w + 2$. Thus, the posterior is proper. \blacksquare

Suppose we use the parameterization $\delta = 1/\lambda$, so that δ is the rate of the Poisson distribution. Then, the Jeffrey's prior is

$$\pi(C, \delta) \propto \left(\frac{e^{1/\delta} \delta - \delta - 1}{\delta^4} \right)^{1/2}$$

and can be verified by transforming back to the original parameterization. Using $d\delta/d\lambda = -\lambda^{-2}$ we obtain

$$\begin{aligned} \pi(C, \lambda) &\propto \left(\frac{e^\lambda \left(\frac{1}{\lambda}\right) - \frac{1}{\lambda} - 1}{\left(\frac{1}{\lambda}\right)^4} \right)^{1/2} \lambda^{-2} \\ &= \left(\frac{e^\lambda - \lambda - 1}{\lambda} \right)^{1/2}. \end{aligned}$$

3.1.2 Geometric Model

Let $f(\lambda|\theta)$ be an exponential distribution parameterized as

$$f(\lambda|\theta) = \frac{1}{\theta} e^{-\lambda/\theta}$$

with $\theta > 0$ and $0 \leq \lambda < \infty$. Then $E[\Lambda] = \theta$ and $Var(\Lambda) = \theta^2$. The exponential-mixed Poisson distribution is

$$\begin{aligned} p_\theta(x) &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \frac{1}{\theta} e^{-\lambda/\theta} d\theta \\ &= \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^x \end{aligned} \quad (3.5)$$

which is also known as the geometric distribution, often used to model the number of failures until the first success when the success probability is $1/(\theta + 1)$.

Consider p_θ as an exponential-mixed Poisson, or geometric, distribution. This means the species' abundances are not all equal. Specifically the distribution of the abundances is exponential. The information matrix can be calculated using (3.2). The $(2, 2)$ element of the information matrix is

$$\begin{aligned} F(C, \theta)_{22} &= -C \left\{ E_X \left[\frac{\partial^2}{\partial \theta^2} \log p_\theta(X) \right] \right\} \\ &= -C \left\{ E_X \left[\frac{\partial^2}{\partial \theta^2} X \log \theta - (X + 1) \log(1 + \theta) \right] \right\} \\ &= -C \left\{ E_X \left[-\frac{X}{\theta^2} + \frac{X + 1}{(1 + \theta)^2} \right] \right\} \\ &= \frac{C}{\theta(1 + \theta)} \end{aligned}$$

where the last equality is true since $E[X] = \theta$. The off-diagonal elements of the information matrix are

$$\begin{aligned} F(C, \theta)_{21} = F(C, \theta)_{12} &= -\frac{\partial}{\partial \theta} \log p_\theta(0) \\ &= -\frac{\partial}{\partial \theta} (-\log(1 + \theta)) \\ &= \frac{1}{1 + \theta}. \end{aligned}$$

The (1, 1) element of the information is

$$\begin{aligned}
F(C, \theta)_{11} &= \frac{1}{C} \frac{1 - p_\theta(0)}{p_\theta(0)} \\
&= \frac{1}{C} \frac{\theta/(1 + \theta)}{1/(1 + \theta)} \\
&= \frac{\theta}{C}.
\end{aligned}$$

The information matrix for C and θ is

$$F(C, \theta) = \begin{pmatrix} \frac{\theta}{C} & \frac{1}{1 + \theta} \\ \frac{1}{1 + \theta} & \frac{1}{C\theta(1 + \theta)} \end{pmatrix} \quad (3.6)$$

and the Jeffreys prior is

$$\pi(C, \theta) \propto \frac{\theta^{1/2}}{\theta + 1}.$$

The posterior using the Jeffreys prior is

$$\begin{aligned}
\pi(C, \theta | \text{data}) &\propto \pi(C, \theta) L(\text{data} | C, \theta) \\
&= \frac{\theta^{1/2}}{\theta + 1} \frac{C!}{(C - w)!} \left(\frac{1}{1 + \theta} \right)^{C - w} \frac{1}{\prod_{j \geq 1} n_j!} \\
&\quad \times \prod_{j \geq 1} \left(\frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta} \right)^j \right)^{n_j} \\
&= \frac{\theta^{1/2}}{1 + \theta} \frac{C!}{(C - w)!} \theta^n (1 + \theta)^{-C - n}.
\end{aligned}$$

Proposition 2. *The posterior for the exponential-mixed Poisson model with Jeffreys prior given as*

$$\pi(C, \theta | \text{data}) \propto \frac{\theta^{1/2}}{1 + \theta} \frac{C!}{(C - w)!} \theta^n (1 + \theta)^{-C - n}$$

is proper.

Proof. Integrating over C and θ we have

$$\begin{aligned}
\int d\pi(C, \theta | \text{data}) &\propto \int_0^\infty \theta^{n-w-1/2} (1+\theta)^{-n} \sum_{C=w}^\infty \frac{C!}{(C-w)!} \left(\frac{1}{1+\theta} \right)^{C-w} \\
&\quad \times \left(\frac{\theta}{1+\theta} \right)^{w+1} d\theta \\
&= \int_0^\infty \theta^{n-w-1/2} (1+\theta)^{-n} d\theta \\
&= \frac{\Gamma(w-1/2)\Gamma(n-w+1/2)}{\Gamma(n)}
\end{aligned}$$

which is finite for any nonempty sample, $n > 0$. ■

3.2 Reference Prior

A corollary from Bernardo and Ramón (1998) can be used to derive the reference prior for the case where our model has one nuisance parameter, $m = 1$. In that case, we can use the fact that the elements of the information matrix in (3.2) will each factor into a function of C and a function of η . Let $S = F^{-1}$ be the covariance matrix. Then,

$$S(C, \eta) = \frac{1}{\det(F(C, \eta))} \begin{pmatrix} C \varrho(\eta) & \frac{\partial}{\partial \eta} \log p_\eta(0) \\ \frac{\partial}{\partial \eta} \log p_\eta(0) & \frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} \end{pmatrix}$$

where $\det(F(C, \eta))$ is a function of η only. Notice that we have the factorization $s_{11}^{-1/2} = a_0(C)b_0(\eta)$ and $f_{22}^{1/2} = a_1(C)b_1(\eta)$, elements of the covariance and information matrices, respectively. We now restate Corollary 1 from Bernardo and Ramón (1998) for the reader's convenience.

Corollary 1. (Bernardo & Ramón 1998): *Suppose the posterior distribution of (ν_1, ν_2) is asymptotically normal with covariance matrix $S(\nu_1, \nu_2)$.*

For ordered parameterization (ν_1, ν_2) ,

1. *if the nuisance parameter space is independent of ν_1 , and*

2. if $s_{11}(\nu_1, \nu_2) = a_0(\nu_1)b_0(\nu_2)$ and $f_{22}(\nu_1, \nu_2) = a_1(\nu_1)b_1(\nu_2)$,

then the reference prior is

$$\pi(\nu_1, \nu_2) = a_0(\nu_1)^{-1/2} b_1(\nu_2)^{1/2}.$$

The nuisance parameter space is independent of the parameter of interest, since for any C we have $\eta \in \mathbb{R}$. The results from Lindsay and Roeder (1987) and Sanathanan (1972) show that the information matrix in (3.2) is the inverse of the covariance matrix for the posterior. Applying the corollary, the reference prior is

$$\pi(C, \eta) \propto C^{-1/2} (\varrho(\eta))^{1/2}.$$

and notice that the jointly derived prior still factorizes into a function of C and a function of η . Moreover, this prior is a product of the two one-dimensional reference priors for C and η .

3.2.1 Poisson Model

In order to find the reference prior for the Poisson model, the information and covariance matrices need to be computed. The information matrix for the Poisson model is given in (3.4). The inverse of the information matrix is

$$S(C, \lambda) = \begin{pmatrix} \frac{C}{e^\lambda - \lambda - 1} & \frac{-\lambda}{e^\lambda - \lambda - 1} \\ \frac{-\lambda}{e^\lambda - \lambda - 1} & \frac{\lambda(e^\lambda - 1)}{C(e^\lambda - \lambda - 1)} \end{pmatrix}$$

and the resulting reference prior is

$$\pi(C, \lambda) \propto C^{-1/2} \lambda^{-1/2}.$$

This prior is improper, yet we will show the posterior is proper.

The posterior is

$$\begin{aligned}
\pi(C, \lambda | \text{data}) &\propto \pi(C, \lambda) L(\text{data} | C, \lambda) \\
&= C^{-1/2} \lambda^{-1/2} \frac{C!}{(C-w)!} (e^{-\lambda})^C \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{1}{j^{n_j}} \right) \lambda^n \\
&\propto C^{-1/2} \frac{C!}{(C-w)!} (e^{-\lambda})^C \lambda^{n-1/2}.
\end{aligned}$$

Proposition 3. *The posterior for the Poisson model using the reference prior is*

$$\pi(C, \lambda | \text{data}) \propto C^{-1/2} \frac{C!}{(C-w)!} (e^{-\lambda})^C \lambda^{n-1/2}$$

and is proper if $n > w$.

Proof. Integrating over C and λ we have

$$\begin{aligned}
\int d\pi(C, \lambda | \text{data}) &\propto \sum_{C=w}^{\infty} C^{-1/2} \frac{C!}{(C-w)!} \int_0^{\infty} \lambda^{n-1/2} e^{-\lambda C} d\lambda \\
&= \sum_{C=w}^{\infty} C^{-1/2} \frac{C!}{(C-w)!} \frac{\Gamma(n+1/2)}{C^{(n+1/2)}} \\
&\propto \sum_{C=w}^{\infty} \frac{C!}{(C-w)!} \frac{1}{C^{(n+1)}} \\
&= \sum_{C=w}^{\infty} \frac{1}{C^{(n-w+1)}} \frac{C-1}{C} \frac{C-2}{C} \dots \frac{C-w+1}{C} \\
&< \sum_{C=w}^{\infty} \frac{1}{C^{(n-w+1)}} \times 1 \\
&< \infty.
\end{aligned}$$

We need $n - w + 1 > 1$ (i.e. $n > w$) for the posterior to be finite. ■

The condition on the sample size in the proof of Proposition 3 requires the number of observed individuals to be greater than the number of observed species in the sample. If only one individual of each species is sampled, then $n = w$. If the sample is empty, then $n = w = 0$; however, an empty sample would not be used for analysis.

3.2.2 Geometric Model

In order to find the reference prior for the geometric model, the information and covariance matrices need to be computed. The information matrix for the geometric model is given in (3.6). The inverse of the information matrix is

$$S(C, \theta) = \begin{pmatrix} \frac{C(1+\theta)}{\theta^2} & -\frac{1+\theta}{\theta} \\ -\frac{1+\theta}{\theta} & \frac{\theta}{(1+\theta)^2} \end{pmatrix} \frac{1}{C}$$

and the resulting reference prior is

$$\pi(C, \theta) \propto C^{-1/2} [\theta(1+\theta)]^{-1/2}.$$

This prior is improper, yet we will show the posterior is proper.

The posterior is

$$\begin{aligned} \pi(C, \lambda | \text{data}) &\propto \pi(C, \lambda) L(\text{data} | C, \lambda) \\ &= C^{-1/2} [\theta(1+\theta)]^{-1/2} \frac{C!}{(C-w)!} \left(\frac{1}{1+\theta} \right)^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \\ &\quad \times \prod_{j \geq 1} \left(\frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^j \right)^{n_j} \\ &\propto C^{-1/2} \frac{C!}{(C-w)!} \theta^{n-1/2} (1+\theta)^{-C-n-1/2}. \end{aligned}$$

Proposition 4. *The posterior for the exponential-mixed Poisson model with reference prior is*

$$\pi(C, \theta | \text{data}) \propto C^{-1/2} \frac{C!}{(C-w)!} \theta^{n-1/2} (1+\theta)^{-C-n-1/2}$$

and is proper.

Proof. Integrating over C and θ we have

$$\begin{aligned}
\int d\pi(C, \theta | \text{data}) &\propto \sum_{C=w}^{\infty} C^{-1/2} \frac{C}{(C-w)!} \int_0^{\infty} \theta^{n-1/2} (1+\theta)^{-C-n-1/2} d\theta \\
&= \sum_{C=w}^{\infty} C^{-1/2} \frac{C}{(C-w)!} \frac{\Gamma(C)\Gamma(n+1/2)}{\Gamma(C+n+1/2)} \\
&\propto \sum_{C=w}^{\infty} C^{-1/2} \frac{C}{(C-w)!} \frac{(C-1)!}{\Gamma(C+n+1/2)} \\
&\propto \sum_{C=w}^{\infty} C^{-1/2} \frac{\overbrace{C(C-1)\cdots(C-w+1)}^{w \text{ terms}}}{\underbrace{(C+n+1/2-1)\cdots(C+n+1/2-C-n)}_{C+n \text{ terms}}} \\
&\quad \times \overbrace{(C-1)(C-2)\cdots(2)(1)}^{C-1 \text{ terms}} \\
&= \sum_{C=w}^{\infty} C^{-1/2} \frac{C(C-1)\cdots(C-w+1)}{(C+n+1/2-1)\cdots(C+n+1/2-n-1)} \\
&\quad \times \frac{C-1}{C-1/2} \frac{C-2}{C-3/2} \cdots \frac{1}{3/2} \frac{1}{1/2} \\
&< \sum_{C=w}^{\infty} \frac{1}{1/2} C^{-1/2} \frac{C(C-1)\cdots(C-w+1)}{(C+n+1/2-1)\cdots(C+n+1/2-n-1)}.
\end{aligned}$$

If $n = w$,

$$\begin{aligned}
\int d\pi(C, \theta | \text{data}) &\propto \sum_{C=w}^{\infty} \frac{1}{1/2} C^{-1/2} \frac{1}{C+w+1/2} \frac{C}{C+w-3/2} \cdots \frac{C-w+1}{C-1/2} \\
&< \sum_{C=w}^{\infty} \frac{1}{1/2} C^{-1/2} \frac{1}{C+w+1/2} \\
&\propto \sum_{C=w}^{\infty} \frac{1}{C^{3/2} + C^{1/2}w + C^{1/2}(1/2)} \\
&< \infty.
\end{aligned}$$

If $n > w$,

$$\int d\pi(C, \theta | \text{data}) < \sum_{C=w}^{\infty} \frac{1}{1/2} C^{-1/2} \frac{1}{C+n+1/2}$$

and the posterior is again finite. ■

3.3 Data Analysis

The Poisson and geometric models fit well for small data sets, so in this section we analyze the data from the Framvaren Fjord, a small data set, where $w = 39$ and $n = 303$.

It is typical for a collection of frequency counts to have a few large abundances. In the Framvaren Fjord data, observed counts for frequencies 165 and 20 are considered as abundant species from a known subpopulation. With this assumption the abundances for the highest frequencies are treated as outliers and not factored into the likelihood. We denote the maximum frequency modeled as τ . The choice of τ is important and model dependent. For our analysis we fix τ for each model, model the frequencies up to τ with a specific abundance distribution, and then the observed frequencies greater than τ are added in later for the final estimate. The value of τ is determined, in part, through a maximum likelihood analysis by calculating a χ^2 goodness-of-fit statistic and choosing the best subset of the data similar to the procedure discussed in Hong et al. (2006). This subsetting problem is further discussed in section 6.3.

Simulation from the posterior for each model uses a Gibbs sampler with the Metropolis and Metropolis-Hastings algorithms. Simulated posteriors contain 500,000 iterations after a 1,000 iteration burn-in period keeping acceptance rates near 30%. We sample alternatively from the full conditional distributions, $\pi(\eta|C, \text{data})$ and $\pi(C|\eta, \text{data})$. The Metropolis algorithm with a normal proposal distribution is used for sampling the nuisance parameter. To sample from the full conditional for C we use direct sampling when possible or a Metropolis-Hastings step when the closed form for the full conditional is not available. In that case, the Metropolis-Hastings step uses a negative binomial proposal distribution.

For the Poisson model using the Jeffreys prior the posterior we are sampling

from is

$$\begin{aligned}\pi(C, \lambda | \text{data}) &\propto \frac{C!}{(C-w)!} e^{-\lambda C} (e^\lambda - \lambda - 1)^{1/2} \lambda^{n-1/2} \\ &\propto \pi(C | \lambda, \text{data}) \pi(\lambda | \text{data})\end{aligned}$$

where

$$\begin{aligned}\pi(\lambda | \text{data}) &\propto \frac{1}{(1 - e^{-\lambda})^{w+1}} (e^\lambda - \lambda - 1)^{1/2} e^{-\lambda w} \lambda^{n-1/2} \\ \pi(C | \lambda, \text{data}) &= \text{negative binomial}(w+1, 1 - e^{-\lambda}).\end{aligned}$$

For the Poisson model using the reference prior, the posterior is

$$\pi(C, \lambda | \text{data}) \propto \frac{C!}{(C-w)!} C^{-1/2} e^{-\lambda C} \lambda^{n-1/2}$$

and the full conditionals are

$$\begin{aligned}\pi(\lambda | C, \text{data}) &\propto e^{-\lambda C} \lambda^{n-1/2} \\ \pi(C | \lambda, \text{data}) &\propto \frac{C!}{(C-w)!} C^{-1/2} e^{-\lambda C}.\end{aligned}$$

For the exponential model using the Jeffreys prior, the posterior is

$$\begin{aligned}\pi(C, \theta | \text{data}) &\propto \frac{C!}{(C-w)!} \theta^{n+1/2} (1 + \theta)^{-C-n-1} \\ &\propto \pi(C | \theta, \text{data}) \pi(\theta | \text{data})\end{aligned}$$

where

$$\begin{aligned}\pi(\theta | \text{data}) &\propto \theta^{n-w-1/2} (1 + \theta)^{-n} \\ \pi(C | \theta, \text{data}) &\propto \text{negative binomial}(w+1, \theta/(1 + \theta)).\end{aligned}$$

For the exponential model using the reference prior, the posterior is

$$\pi(C, \theta | \text{data}) \propto \frac{C!}{(C-w)!} C^{-1/2} \theta^{n-1/2} (1 + \theta)^{-C-n-1/2}$$

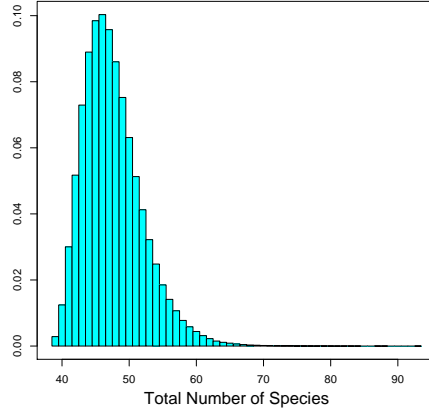
and the full conditionals are

$$\begin{aligned}\pi(\theta|C, \text{data}) &\propto \theta^{n-1/2}(1+\theta)^{-C-n-1/2} \\ \pi(C|\theta, \text{data}) &\propto \frac{C!}{(C-w)!}C^{-1/2}(1+\theta)^{-C-n-1/2}.\end{aligned}$$

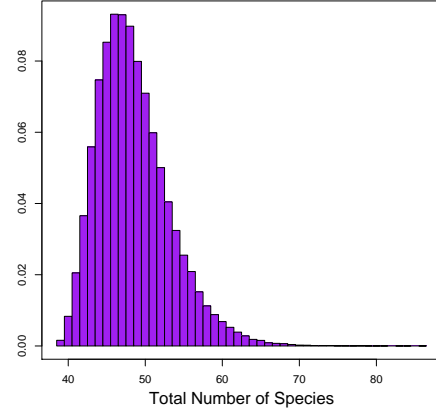
Figure 3.1 shows posterior simulations for C from each of the two models and each of the two priors for the Framvaren Fjord data. The value for τ is set at 5 for the Poisson model and 12 for the geometric model. Diagnostic plots for our samplers are shown in Figures 3.2 and 3.3. For each model-prior combination we show trace plots and autocorrelation plots for C . The diagnostic plots for each model show that we are confident our MCMC has converged.

Bayesian estimates for C are shown in Tables 3.1 and 3.2. The median of the posterior sample is considered as a point estimate and a 95% central credible interval is constructed using the 0.025 and 0.975 quantiles of the posterior. The central credible interval gives a range of values with the probability of being below or above the interval equal to 0.025. Conditional maximum likelihood estimates for this problem are shown for comparison. Frequentist estimates are computed as in Hong et al. (2006). Confidence intervals for the MLEs are log transformed confidence intervals based on asymptotic normality (Chao 1987). The log tranformed confidence intervals will be used throughtout the dissertation. However, credible intervals can also be compared to profile likelihood intervals (Cormack 1992). For example, the geometric model (Table 3.2) with maximum likelihood point estimate of 58.34 has an associated profile likelihood interval of (46.89, 73.82).

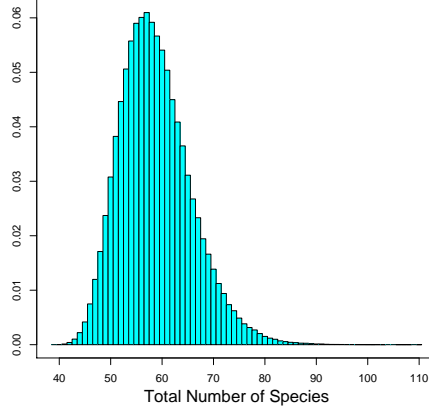
We can see that the Bayesian estimates are similar to the maximum likelihood estimates for each model. Specifically, the confidence intervals are comparable with the credible interval estimates and the posterior median estimates for C are similar to the maximum likelihood estimates.



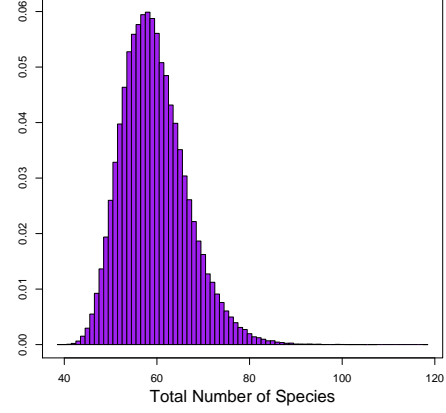
(a) Poisson-Jeffreys



(b) Poisson-reference



(c) Geometric-Jeffreys

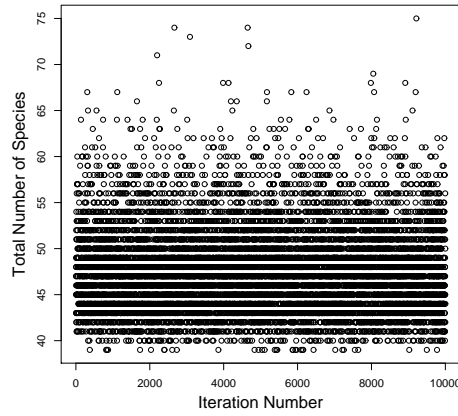


(d) Geometric-reference

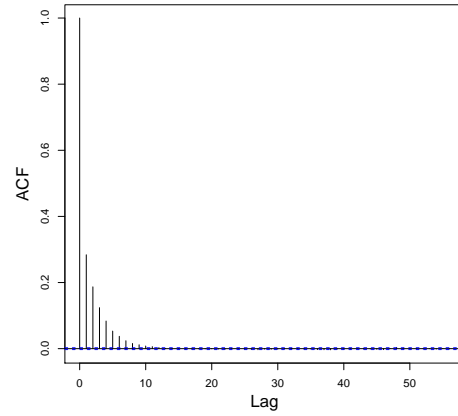
Figure 3.1: Histograms of posterior samples from $\pi(C|data)$ for (a) Poisson model with Jeffreys prior, (b) Poisson model with reference prior, (c) geomtric model with Jeffreys prior, and (d) geometric model with reference prior.

Table 3.1: Estimates for the Poisson model. Bayesian estimates resulting from the Jeffreys and reference priors are shown as well as the maximum likelihood estimate (MLE).

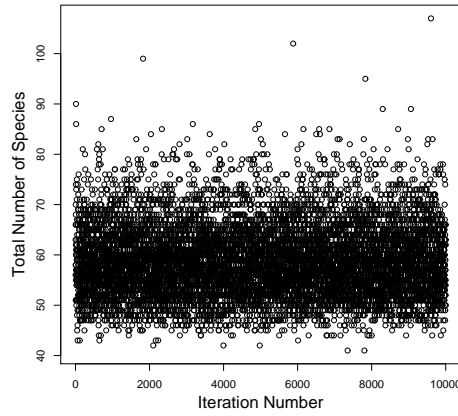
Model	Point estimate	CI
Jeffreys	47	(41,58)
Reference	48	(41,60)
MLE	47.35	(42.23,60.60)



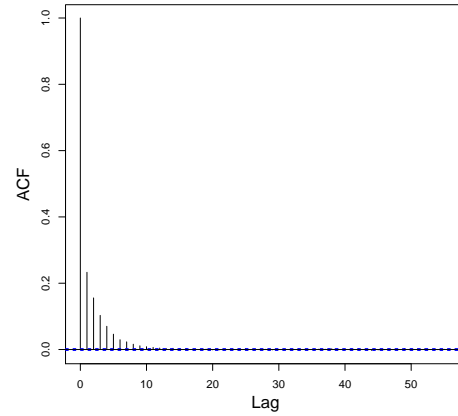
(a) Trace plot in Poisson model



(b) Autocorrelation plot in Poisson model



(c) Trace plot in geometric model

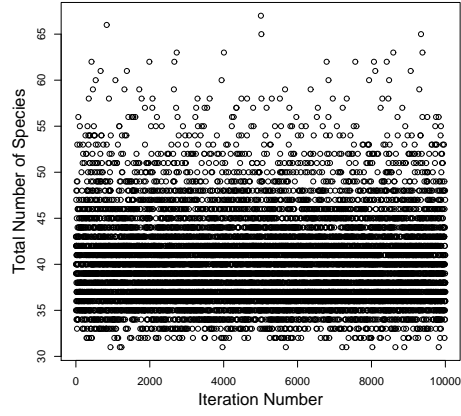


(d) Autocorrelation plot in geometric model

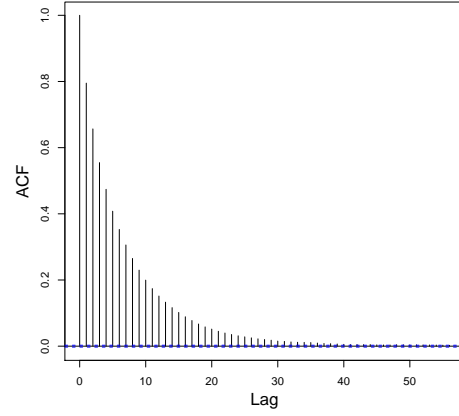
Figure 3.2: Diagnostic plots for models using Jeffreys priors.

Table 3.2: Estimates for the geometric model. Bayesian estimates resulting from the Jeffreys and reference priors are shown as well as the maximum likelihood estimate (MLE).

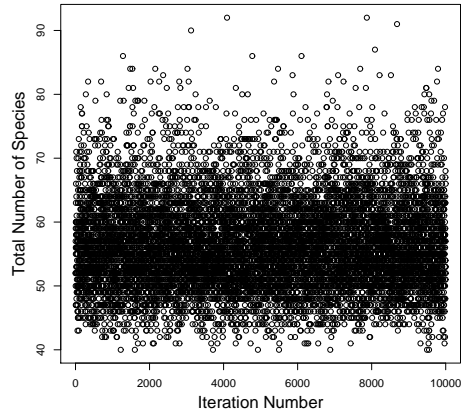
Model	Point estimate	CI
Jeffreys	58	(47,74)
Reference	59	(48,75)
MLE	58.34	(49.40,74.99)



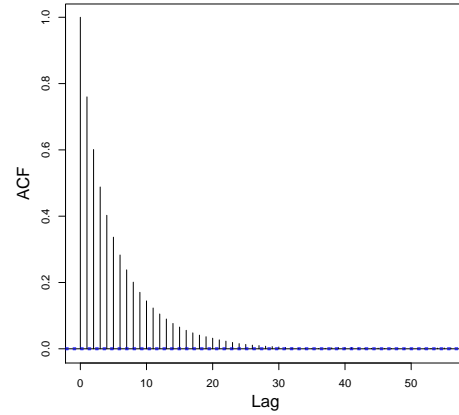
(a) Trace plot in Poisson model



(b) Autocorrelation plot in Poisson model



(c) Trace plot in geometric model



(d) Autocorrelation plot in geometric model

Figure 3.3: Diagnostic plots for models using reference priors.

We assess the fit of each model using plots of the fitted values. Figure 3.4 shows plots of the raw data with expected values for the frequencies n_1, n_2, \dots, n_τ using the median posterior values from the reference priors. We see that for this small data set the geometric model's fit is acceptable.

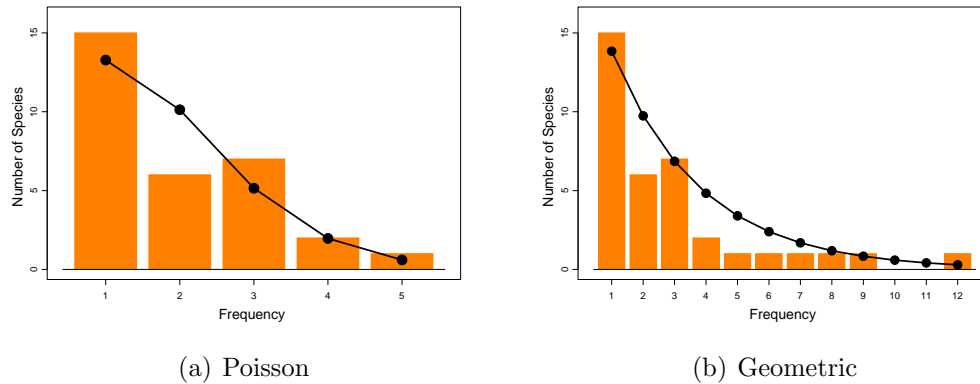


Figure 3.4: Expected frequencies for Poisson and geometric models using reference priors.

The Jeffreys and reference priors for geometric model give similar results. The choice between these two priors minimally affects the resulting estimates. However, model selection is important in this problem and choice of model can highly influence the final estimate.

Chapter 4

Two Nuisance Parameters

We now extend the methods of deriving objective priors for the case where the parameter, η , describing the abundance distribution is a two-dimensional nuisance parameter. We examine in detail the case when p_η is a negative binomial. This two parameter model is an extension of the geometric model discussed in Chapter 3. The model now has a total of three parameters which we notate as C , η_1 , and η_2 . We will refer to the likelihood in (1.1) and the information matrix in (1.6). For two-dimensional η we have

$$F(C, \eta) = \begin{pmatrix} \frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} & -\frac{\partial}{\partial \eta_1} \log p_\eta(0) & -\frac{\partial}{\partial \eta_2} \log p_\eta(0) \\ -\frac{\partial}{\partial \eta_1} \log p_\eta(0) & C \varrho(\eta)_{11} & C \varrho(\eta)_{12} \\ -\frac{\partial}{\partial \eta_2} \log p_\eta(0) & C \varrho(\eta)_{21} & C \varrho(\eta)_{22} \end{pmatrix} \quad (4.1)$$

with $\varrho(\eta)_{kl} = -E_X \left[\frac{\partial^2}{\partial \eta_k \partial \eta_l} \log p_\eta(X) \right]$ for $k, l = 1, 2$, where we use the simplifying assumption

$$E_X \left[\left(\frac{\partial}{\partial \eta_1} \log p_\eta(X) \right) \left(\frac{\partial}{\partial \eta_2} \log p_\eta(X) \right) \right] = - \left\{ E_X \left[\frac{\partial^2}{\partial \eta_1 \partial \eta_2} \log p_\eta(X) \right] \right\}.$$

4.1 Jeffreys Prior

In this section the prior for (C, η_1, η_2) will be derived using the multivariate Jeffreys' rule described in section 1.3.2 replacing the Fisher information with the

information in (4.1).

The determinant of the information matrix in (4.1) is

$$\begin{aligned}\det(F(C, \eta)) &= F(C, \eta)_{11}D_1 - F(C, \eta)_{12}D_2 + F(C, \eta)_{13}D_3 \\ &= \frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} D_1 - \frac{\partial}{\partial \eta_1} \log p_\eta(0) D_2 - \frac{\partial}{\partial \eta_2} \log p_\eta(0) D_3\end{aligned}$$

where D_1 , D_2 , and D_3 are the minors of elements $F(C, \eta)_{11}$, $F(C, \eta)_{12}$, and $F(C, \eta)_{13}$, respectively. The minors are

$$\begin{aligned}D_1 &= \begin{vmatrix} F(C, \eta)_{22} & F(C, \eta)_{23} \\ F(C, \eta)_{32} & F(C, \eta)_{33} \end{vmatrix} \\ &= C^2 \left(\varrho(\eta)_{11} \varrho(\eta)_{22} - \varrho(\eta)_{12} \varrho(\eta)_{21} \right), \\ D_2 &= \begin{vmatrix} F(C, \eta)_{21} & F(C, \eta)_{23} \\ F(C, \eta)_{31} & F(C, \eta)_{33} \end{vmatrix} \\ &= C \left(- \left(\frac{\partial}{\partial \eta_1} \log p_\eta(0) \right) \varrho(\eta)_{22} + \left(\frac{\partial}{\partial \eta_2} \log p_\eta(0) \right) \varrho(\eta)_{12} \right),\end{aligned}$$

and

$$\begin{aligned}D_3 &= \begin{vmatrix} F(C, \eta)_{21} & F(C, \eta)_{22} \\ F(C, \eta)_{31} & F(C, \eta)_{32} \end{vmatrix} \\ &= C \left(- \left(\frac{\partial}{\partial \eta_1} \log p_\eta(0) \right) \varrho(\eta)_{12} + \left(\frac{\partial}{\partial \eta_2} \log p_\eta(0) \right) \varrho(\eta)_{11} \right).\end{aligned}$$

The Jeffreys prior is the square root of this determinant and is of the form

$$\pi(C, \eta) \propto C^{1/2} \pi(\eta). \quad (4.2)$$

Jeffreys prior for C and η factors into two independent priors. Since $\pi(C, \eta)$ is an improper prior (although $\pi(\eta)$ may be integrable over values of η), we want to show the posterior is proper. Using $\pi(C) = C$ as an upper bound for $\pi(C) = C^{1/2}$, the results follows similarly from Chapter 3; i.e., an upper bound to the marginal posterior for η is

$$\pi(\eta|\text{data}) = \pi(\eta) \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \frac{1}{(1 - p_\eta(0))^{w+1}}.$$

When considering a model after specifying p_η , integrability of the posterior can be shown if $\int \pi(\eta|\text{data})d\eta < \infty$, where $\eta = (\eta_1, \eta_2)$.

4.2 Reference Prior

First, let us look at the inverse of the information matrix since this covariance matrix is needed in the derivation of the reference prior. The information matrix with two nuisance parameters is given in (4.1). Notice the lower right block of this matrix contains the Fisher information of the nuisance parameters multiplied by C . The inverse is

$$S(C, \eta_1, \eta_2) = \begin{pmatrix} Cs_{11}(\eta) & s_{12}(\eta) & s_{13}(\eta) \\ s_{21}(\eta) & \frac{1}{C}s_{22}(\eta) & \frac{1}{C}s_{23}(\eta) \\ s_{31}(\eta) & \frac{1}{C}s_{32}(\eta) & \frac{1}{C}s_{33}(\eta) \end{pmatrix}$$

where $s_{ij}(\eta)$ is a function only of η for row i and column j . When we take the inverse of this matrix, there is no guarantee that the elements corresponding to the nuisance parameters will factor into a function for η_1 and for η_2 . However, the elements in the information matrix will factor into a function of C and $\eta = (\eta_1, \eta_2)$.

Proposition 3 in Bernardo and Ramón (1998) shows a general multivariate method to derive reference priors. We use this procedure to determine the reference prior when $m = 2$. The (2×2) upper matrix of $S(C, \eta_1, \eta_2)$ is

$$S_2(C, \eta_1, \eta_2) = \begin{pmatrix} Cs_{11}(\eta) & s_{12}(\eta) \\ s_{21}(\eta) & \frac{1}{C}s_{22}(\eta) \end{pmatrix}$$

with inverse matrix

$$H_2(C, \eta_1, \eta_2) = \begin{pmatrix} \frac{1}{C}h_{11}(\eta) & h_{12}(\eta) \\ h_{21}(\eta) & Ch_{22}(\eta) \end{pmatrix}.$$

Using the ordered parameterization (C, η_1, η_2) , the conditional reference prior for η_2 is

$$\pi(\eta_2|C, \eta_1) \propto \varrho(\eta)_{22}^{1/2}.$$

The conditional reference prior for η_1 is

$$\begin{aligned}
\pi(\eta_1|C) &\propto \exp \left[\int \log(C^{1/2} h_{22}^{1/2}) \pi(\eta_2|C, \eta_1) d\eta_2 \right] \\
&= \exp \left[\int \left(\log C^{1/2} + \log h_{22}^{1/2} \right) \pi(\eta_2|C, \eta_1) d\eta_2 \right] \\
&= \exp \left[\log C^{1/2} \int \pi(\eta_2|C, \eta_1) d\eta_2 \right] \\
&\quad \times \exp \left[\int \left(\log h_{22}^{1/2} \right) \pi(\eta_2|C, \eta_1) d\eta_2 \right] \tag{4.3}
\end{aligned}$$

$$\propto \exp \left[\int \left(\log h_{22}^{1/2} \right) \pi(\eta_2|C, \eta_1) d\eta_2 \right] \tag{4.4}$$

If the conditional reference prior for η_2 is not proper, than a compact approximation is required for the corresponding integral. In (4.3) we are able to consider the first exponential as a constant with respect to η_1 since $\int \pi(\eta_2|C, \eta_1) d\eta_2 = 1$ if the conditional priors is proper or if a compact approximation is used.

The marginal reference prior of C is

$$\begin{aligned}
\pi(C) &\propto \exp \left[\int \int \log(C^{-1/2} g_{11}^{-1/2}) \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) d\eta_2 d\eta_1 \right] \\
&= \exp \left[\int \int \left(\log(C^{-1/2}) \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) \right. \right. \\
&\quad \left. \left. + \log(g_{11}^{-1/2}) \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) \right) d\eta_2 d\eta_1 \right] \\
&= \exp \left[\log(C^{-1/2}) \int \int \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) d\eta_2 d\eta_1 \right] \\
&\quad \times \exp \left[\int \int \log(g_{11}^{-1/2}) \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) d\eta_2 d\eta_1 \right] \\
&\propto \exp \left[\log(C^{-1/2}) \int \int \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) d\eta_2 d\eta_1 \right] \\
&= C^{-1/2}
\end{aligned}$$

where the last equality is true when $\int \int \pi(\eta_2|C, \eta_1) \pi(\eta_1|C) d\eta_2 d\eta_1 = 1$. This can be shown directly if the priors on η are proper. If either $\pi(\eta_2|C, \eta_1)$ or $\pi(\eta_1|C)$ is improper, a compact approximation is required to show the result.

The joint reference prior for (C, η) is

$$\pi(C, \eta) \propto C^{-1/2} \pi(\eta)$$

where $\pi(\eta)$ is the reference prior for a likelihood of C i.i.d. replicates from p_η .

4.3 Negative Binomial Model

We will now examine the negative binomial model. The form of the Jeffreys and reference priors will be derived. The problem becomes very complex and it is difficult to implement the procedure as in Chapter 3. Instead, we use simplified versions of the priors and show conditions in order to have a proper posterior.

Let $f(\lambda|\alpha, \beta)$ be a gamma distribution parameterized as

$$f(\lambda|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-\lambda/\beta} \lambda^{\alpha-1}$$

with $\alpha, \beta > 0$ and $0 \leq \lambda < \infty$. Then $E[\lambda] = \alpha\beta$ and $Var(\lambda) = \alpha\beta^2$. A special case is the exponential distribution when $\alpha = 1$. The gamma mixed Poisson distribution is

$$\begin{aligned} p_{\alpha, \beta}(x) &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-\lambda/\beta} \lambda^{\alpha-1} d\lambda \\ &= \frac{1}{x! \Gamma(\alpha) \beta^\alpha} \int_0^\infty e^{-\lambda/(\frac{\beta}{\beta+1})} \lambda^{(x+\alpha)-1} d\lambda \\ &= \frac{\Gamma(x+\alpha) (\frac{\beta}{\beta+1})^{x+\alpha}}{x! \Gamma(\alpha) \beta^\alpha} \\ &= \frac{\Gamma(x+\alpha)}{x! \Gamma(\alpha)} \left(\frac{\beta}{\beta+1} \right)^x \left(\frac{1}{\beta+1} \right)^\alpha \end{aligned} \tag{4.5}$$

which is also known as the negative binomial distribution, often used to model the number of failures until α successes with success probability $1/(\beta+1)$. We have $E[X] = \alpha\beta$.

The likelihood for this model is

$$\begin{aligned}
L(\text{data}|C, \eta) &= \frac{C!}{(C-w)!} \left(\frac{1}{\beta+1} \right)^{\alpha(C-w)} \frac{1}{\prod_{j \geq 1} n_j!} \\
&\quad \times \prod_{j \geq 1} \left(\frac{\Gamma(j+\alpha)}{j! \Gamma(\alpha)} \left(\frac{\beta}{\beta+1} \right)^j \left(\frac{1}{\beta+1} \right)^\alpha \right)^{n_j} \\
&= \frac{C!}{(C-w)!} (\beta+1)^{-\alpha C-n} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \Gamma(j+\alpha)^{n_j} \\
&\quad \times \prod_{j \geq 1} \left(\frac{1}{j!} \right)^{n_j} \Gamma(\alpha)^{-w} \beta^n. \tag{4.6}
\end{aligned}$$

Computing the 3×3 information matrix entails computing each element in (4.1). For the (1,1) element of the information we have

$$F(C, \alpha, \beta)_{11} = \frac{1}{C} \frac{1 - (\beta+1)^{-\alpha}}{(\beta+1)^{-\alpha}}.$$

For the (1,2) element of the information we have

$$F(C, \alpha, \beta)_{12} = -\frac{\partial}{\partial \alpha} (-\alpha \log(\beta+1)) = \log(\beta+1).$$

For the (1,3) element of the information we have

$$F(C, \alpha, \beta)_{13} = -\frac{\partial}{\partial \beta} (-\alpha \log(\beta+1)) = \frac{\alpha}{\beta+1}.$$

For the (2,2) element of the information we have

$$\begin{aligned}
F(C, \alpha, \beta)_{22} &= -C \left\{ E \left[\frac{\partial^2}{\partial \alpha^2} \log p_{\alpha\beta}(x) \right] \right\} \\
&= -C \left\{ E \left[\frac{\partial^2}{\partial \alpha^2} \log \Gamma(x+\alpha) - \log x! - \log \Gamma(\alpha) + x \log \beta \right. \right. \\
&\quad \left. \left. - (x+\alpha) \log(\beta+1) \right] \right\} \\
&= -C \left\{ E \left[\frac{\partial}{\partial \alpha} \psi(x+\alpha) - \psi(\alpha) - \log(\beta+1) \right] \right\} \\
&= -C \{ E [\psi_1(x+\alpha) - \psi_1(\alpha)] \} \\
&= C (\psi_1(\alpha) - E [\psi_1(x+\alpha)])
\end{aligned}$$

where ψ is the derivative of the log of the gamma function, known as the digamma function, and ψ_1 is the derivative of ψ . For the (2, 3) element of the information we have

$$\begin{aligned}
F(C, \alpha, \beta)_{23} &= -C \left\{ E \left[\frac{\partial^2}{\partial \alpha \partial \beta} \log p_{\alpha\beta}(x) \right] \right\} \\
&= -C \left\{ E \left[\frac{\partial}{\partial \alpha} \left(\frac{x}{\beta} - \frac{x + \alpha}{\beta + 1} \right) \right] \right\} \\
&= -C \left\{ E \left[-\frac{1}{\beta + 1} \right] \right\} \\
&= \frac{C}{\beta + 1}.
\end{aligned}$$

Finally, for the (3, 3) element of the information we have

$$\begin{aligned}
F(C, \alpha, \beta)_{33} &= -C \left\{ E \left[\frac{\partial^2}{\partial \beta^2} \log p_{\alpha\beta}(x) \right] \right\} \\
&= -C \left\{ E \left[-\frac{x}{\beta^2} + \frac{x + \alpha}{(\beta + 1)^2} \right] \right\} \\
&= \frac{\alpha C}{\beta(\beta + 1)}.
\end{aligned}$$

Hence, the information for C , α and β is

$$F(C, \alpha, \beta) = \begin{pmatrix} \frac{1 - (\beta + 1)^{-\alpha}}{C(\beta + 1)^{-\alpha}} & \ln(\beta + 1) & \frac{\alpha}{\beta + 1} \\ \ln(\beta + 1) & C(\Psi_1(\alpha) - E_X[\Psi_1(X + \alpha)]) & \frac{C}{\beta + 1} \\ \frac{\alpha}{\beta + 1} & \frac{C}{\beta + 1} & \frac{\alpha C}{\beta(\beta + 1)} \end{pmatrix}.$$

The elements of the information matrix factor into a function of C and the nuisance parameters, but notice that the elements of the matrix do not additionally factor into functions of α and β . In particular, note the (2, 2) element is a function of α and β through the expectation.

4.3.1 Jeffreys Prior

Taking the square root of the determinant of the information, Jeffreys prior is

$$\begin{aligned}\pi(C, \alpha, \beta) &\propto [\det I(C, \alpha, \beta)]^{1/2} \\ &= C^{1/2} \left[\frac{1 - (\beta + 1)^{-\alpha}}{(\beta + 1)^{-\alpha+1}} \frac{\alpha}{\beta} r + \frac{2\alpha \ln(\beta + 1)}{(\beta + 1)^2} \right. \\ &\quad \left. - \frac{\alpha^2 r}{(\beta + 1)^2} - \frac{1 - (\beta + 1)^{-\alpha}}{(\beta + 1)^{-\alpha+1}} - \frac{\alpha(\ln(\beta + 1))^2}{\beta(\beta + 1)} \right]^{1/2}\end{aligned}$$

where

$$r = C (\Psi_1(\alpha) - E_X [\Psi_1(X + \alpha)]).$$

Therefore, the form of the Jeffreys prior is

$$\pi(C, \alpha, \beta) \propto C^{1/2} \pi(\alpha, \beta)$$

which is again a product of two independent priors.

The posterior is

$$\begin{aligned}\pi(C, \alpha, \beta | \text{data}) &\propto \pi(C, \alpha, \beta) L(\text{data} | C, \alpha, \beta) \\ &= C^{1/2} \pi_J(\alpha, \beta) \frac{C!}{(C - w)!} (\beta + 1)^{-\alpha C - n} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \Gamma(j + \alpha)^{n_j} \\ &\quad \times \prod_{j \geq 1} \left(\frac{1}{j!} \right)^{n_j} \Gamma(\alpha)^{-w} \beta^n.\end{aligned}$$

The joint Jeffreys prior is complicated in this three parameter model, so we propose a simplification of the prior by taking the product of the diagonal elements instead of the determinant, obtaining

$$\pi_D(C, \alpha, \beta) \propto C^{1/2} \left(\frac{1 - (\beta + 1)^{-\alpha}}{(\beta + 1)^{-\alpha}} (\Psi_1(\alpha) - E_X [\Psi_1(X + \alpha)]) \frac{\alpha}{\beta(\beta + 1)} \right)^{1/2}.$$

Notice this preserves the marginal prior for C and the factorization of the prior.

Also, $\pi_D(C, \alpha, \beta) > \pi(C, \alpha, \beta)$.

The next step is to show this prior yields a proper posterior. An upper bound to the posterior for (C, α, β) is

$$\begin{aligned} \pi_D(C, \alpha, \beta | \text{data}) &= \pi_D(\alpha, \beta) \frac{C!}{(C-w)!} (\beta+1)^{-\alpha C-n} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \Gamma(j+\alpha)^{n_j} \\ &\times \prod_{j \geq 1} \left(\frac{1}{j!} \right)^{n_j} \Gamma(\alpha)^{-w} \beta^n (1 - (\beta+1)^{-\alpha})^{-w-1}. \end{aligned}$$

To prove the posterior is proper, $\pi_D(C, \alpha, \beta | \text{data})$ must be shown to be integrable. Using two different parameterizations of the negative binomial, including an orthogonal parameterization and trying several different upper bounds on $\pi_D(C, \alpha, \beta | \text{data})$, does not make this problem tractable.

4.3.2 Reference Prior

Using (4.5) the reference prior is

$$\pi(C, \alpha, \beta) \propto C^{-1/2} \pi(\alpha, \beta).$$

where $\pi(\alpha, \beta)$ is the reference prior for C i.i.d. replicates from a negative binomial distribution. The conditional reference prior for β is

$$\pi(\beta) \propto \beta^{-1/2} (1 + \beta)^{-1/2}$$

and the conditional reference prior for α can be found using (4.4) and requires inverting the information matrix. Also, a compact approximation for α and β must be used since the priors are improper.

4.4 Data Analysis

The negative binomial distribution is used to model the abundances of the Framvaren Fjord and Lepidoptera data. With this model's additional parameter, the negative binomial is a more flexible model and can provide a better fit to the data.

Orthogonal parameterization for the negative binomial distribution (Huzar-bazar 1950) is used in the implementation. The Gibbs sampler converges faster when the orthogonal parameterization is used compared to the description given previously in (4.5). The parameterization is

$$p_{\theta_1, \theta_2}(x) = \frac{\Gamma(1 + \theta_1 + x)}{\Gamma(1 + \theta_1)\Gamma(x + 1)} \left(\frac{1 + \theta_1}{1 + \theta_1 + \theta_2} \right)^{1 + \theta_1} \left(\frac{\theta_2}{1 + \theta_1 + \theta_2} \right)^x$$

where $\theta_1 > -1$, $\theta_2 > 0$, $x = 0, 1, \dots$, and $E[X] = \theta_2$.

In order to simplify the form of the prior, we recommend the use of the marginal prior for the number of species and a marginal prior for the nuisance parameters. In this implementation we chose independent Cauchy priors with scale parameter equal to one, with density function

$$f(x) = \frac{2}{\pi(1 + x^2)}$$

for $x > 0$. The priors are

$$\pi(\theta_1) = \frac{2}{\pi(1 + (\theta_1 + 1)^2)} = \frac{2}{\pi(\theta_1^2 + 2\theta_1 + 2)}$$

and

$$\pi(\theta_2) = \frac{2}{\pi(1 + \theta_2^2)}.$$

There are many choices for noninformative priors on the nuisance parameters. We suggest using a prior that would be suitable for C independent replicates from the abundance distribution.

Simulation from the posterior for each model uses a Gibbs sampler with the Metropolis and Metropolis-Hastings algorithms. We sample alternatively from the full conditional distributions, $\pi(\theta_1, \theta_2 | C, \text{data})$ and $\pi(C | \theta_1, \theta_2, \text{data})$. The Metropolis algorithm with a bivariate normal proposal distribution with correlation equal to zero (for independent draws) is used for sampling the nuisance parameter. To sample from the full conditional for C we use a Metropolis-Hastings

step with a negative binomial proposal distribution. There is high correlation between samples from the full conditional for C and the full conditional for the nuisance parameters (Wang et al. 2007). In order to achieve convergence in our sampler, diagnostic plots must be examined carefully.

For the negative binomial model using the Jeffreys prior, $\pi(C) \propto C^{1/2}$, the posterior we are sampling from is

$$\begin{aligned}
\pi(C, \theta_1, \theta_2 | \text{data}) &\propto \pi(C, \theta_1, \theta_2) L(\text{data} | C, \theta_1, \theta_2) \\
&\propto C^{1/2} (\theta_1^2 + 2\theta_1 + 2)^{-1} (1 + \theta_2^2)^{-1} \frac{C!}{(C-w)!} \frac{1}{\prod_{j \geq 1} n_j!} \\
&\quad \times \left(\frac{1 + \theta_1}{1 + \theta_1 + \theta_2} \right)^{(1+\theta_1)(C-w)} \prod_{j \geq 1} \left[\frac{\Gamma(1 + \theta_1 + x)}{\Gamma(1 + \theta_1) \Gamma(x + 1)} \right. \\
&\quad \times \left. \left(\frac{1 + \theta_1}{1 + \theta_1 + \theta_2} \right)^{(1+\theta_1)} \left(\frac{\theta_2}{1 + \theta_1 + \theta_2} \right)^x \right]^{n_j} \\
&= C^{1/2} \frac{C!}{(C-w)!} (1 + \theta_1 + \theta_2)^{-C(1+\theta_1)-n} (1 + \theta_1)^{C(1+\theta_1)} \\
&\quad \times (\theta_1^2 + 2\theta_1 + 2)^{-1} (1 + \theta_2^2)^{-1} \theta_2^n \prod_{j \geq 1} \Gamma(1 + \theta_1 + j)^{n_j} \\
&\quad \times \Gamma(1 + \theta_1)^{-w}
\end{aligned}$$

and the full conditionals are

$$\begin{aligned}
\pi(\theta_1, \theta_2 | C, \text{data}) &\propto (1 + \theta_1 + \theta_2)^{-C(1+\theta_1)-n} (1 + \theta_1)^{C(1+\theta_1)} (1 + \theta_2^2)^{-1} \\
&\quad \times (\theta_1^2 + 2\theta_1 + 2)^{-1} \theta_2^n \prod_{j \geq 1} \Gamma(1 + \theta_1 + j)^{n_j} \times \Gamma(1 + \theta_1)^{-w}
\end{aligned}$$

and

$$\begin{aligned}
\pi(C | \theta_1, \theta_2, \text{data}) &\propto C^{1/2} \frac{C!}{(C-w)!} (1 + \theta_1 + \theta_2)^{-C(1+\theta_1)-n} (1 + \theta_1)^{C(1+\theta_1)} \\
&= C^{1/2} \frac{C!}{(C-w)!} \left(\frac{1 + \theta_1}{1 + \theta_1 + \theta_2} \right)^{C(1+\theta_1)}
\end{aligned}$$

where we block together θ_1 and θ_2 and simultaneously sample (θ_1, θ_2) from their joint distribution.

For the negative binomial model using the reference prior, $\pi(C) \propto C^{-1/2}$, the posterior we are sampling from is

$$\begin{aligned}\pi(C, \theta_1, \theta_2 | \text{data}) &\propto C^{-1/2} \frac{C!}{(C-w)!} (1 + \theta_1 + \theta_2)^{-C(1+\theta_1)-n} (1 + \theta_1)^{C(1+\theta_1)} \\ &\quad \times (\theta_1^2 + 2\theta_1 + 2)^{-1} (1 + \theta_2^2)^{-1} \theta_2^n \prod_{j \geq 1} \Gamma(1 + \theta_1 + j)^{n_j} \\ &\quad \times \Gamma(1 + \theta_1)^{-w}\end{aligned}$$

and the full conditionals are

$$\begin{aligned}\pi(\theta_1, \theta_2 | C, \text{data}) &\propto (1 + \theta_1 + \theta_2)^{-C(1+\theta_1)-n} (1 + \theta_1)^{C(1+\theta_1)} (1 + \theta_2^2)^{-1} \\ &\quad \times (\theta_1^2 + 2\theta_1 + 2)^{-1} \theta_2^n \prod_{j \geq 1} \Gamma(1 + \theta_1 + j)^{n_j} \times \Gamma(1 + \theta_1)^{-w}\end{aligned}$$

and

$$\pi(C | \theta_1, \theta_2, \text{data}) \propto C^{-1/2} \frac{C!}{(C-w)!} \left(\frac{1 + \theta_1}{1 + \theta_1 + \theta_2} \right)^{C(1+\theta_1)}.$$

4.4.1 Framvaren Fjord

Simulated posteriors contain 1,000,000 iterations after a 1,000 iteration burn-in period keeping acceptance rates near 30%. Figure 4.1 shows posterior simulations for C from each of the two priors. The shape of the distributions are similar. The model based on the Jeffreys prior has a heavier tail. The value for τ is set at 20. Diagnostic plots for our samplers are shown in Figures 4.2. For each prior we show trace plots and autocorrelation plots for C . Despite the expected high autocorrelations, we have run enough simulations for a large effective sample size. The trace plots indicate good mixing of the sampler. The diagnostic plots for each model show that we are confident our MCMC has converged.

Bayesian estimates for C are shown in Table 4.1. The median of the posterior sample is considered as a point estimate and a 95% central credible interval is constructed using the 0.025 and 0.975 quantiles of the posterior. Conditional max-

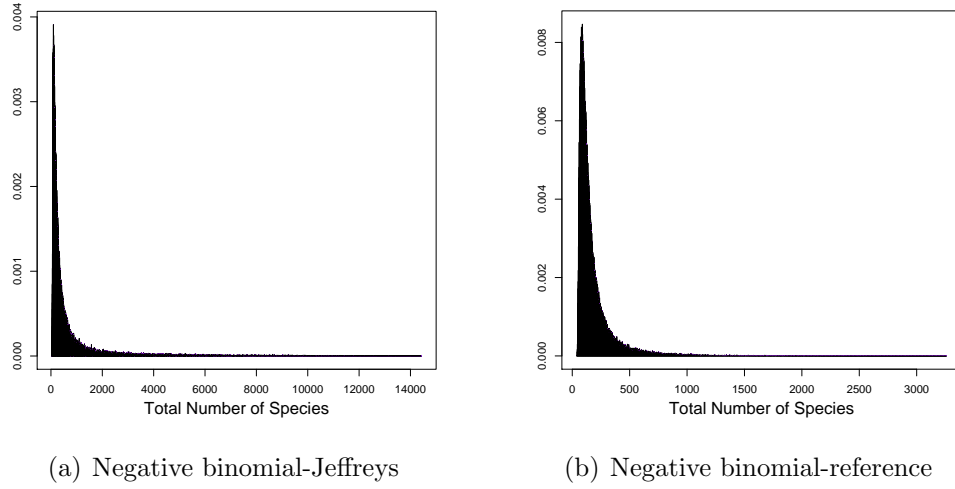


Figure 4.1: Histograms of posterior samples from $\pi(C|\text{data})$. (a) Negative binomial model with Jeffreys prior. (b) Negative binomial model with reference prior.

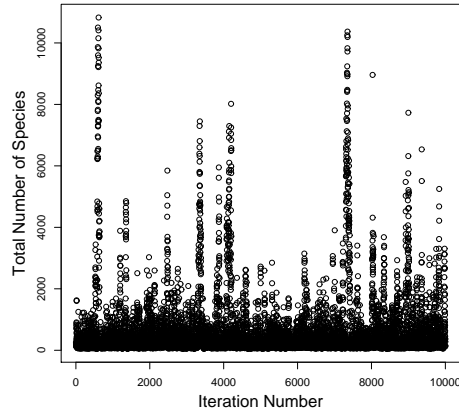
Table 4.1: Estimates for the negative binomial model. Bayesian estimates resulting from the Jeffreys and reference priors are shown as well as the maximum likelihood estimate (MLE).

Model	Point estimate	CI
Jeffreys	248	(70,3864)
Reference	126	(58,659)
MLE	1728.5	(49.35,275784.40)

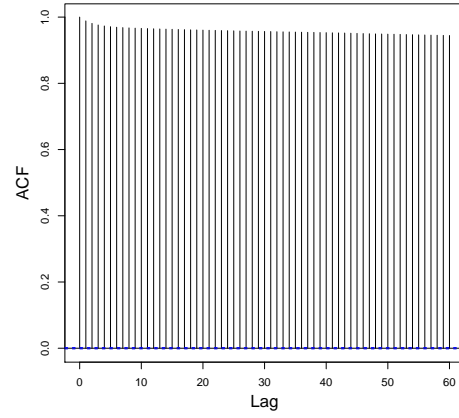
imum likelihood estimates and log transformed confidence intervals are shown.

We can see that the Bayesian estimates are much smaller than the maximum likelihood estimates for each model. In fact, the MLE is unstable for this data set, indicated by the large variance of the MLE, making the estimates unreliable. The point estimate for the Jeffreys prior is larger than the point estimate for the reference prior.

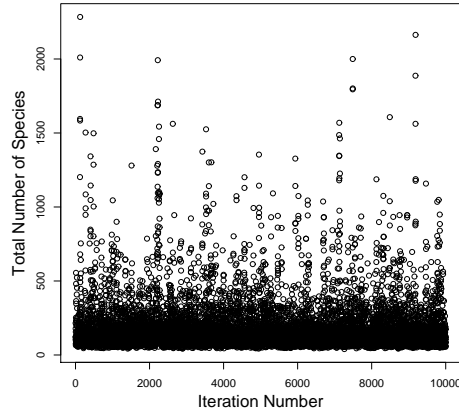
We assess the fit of each model using plots of the fitted values. Figure 4.3 shows plots of the raw data with expected values for the frequencies n_1, n_2, \dots, n_τ



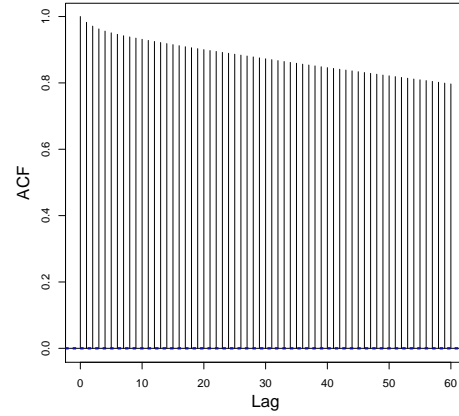
(a) Trace plot for negative binomial-Jeffreys



(b) Autocorrelation plot for negative binomial-Jeffreys



(c) Trace plot for negative binomial-reference



(d) Autocorrelation plot for negative binomial-reference

Figure 4.2: Diagnostic plots for the negative binomial model.

using the median posterior values. We see that for this data set the negative binomial model's fit is acceptable.

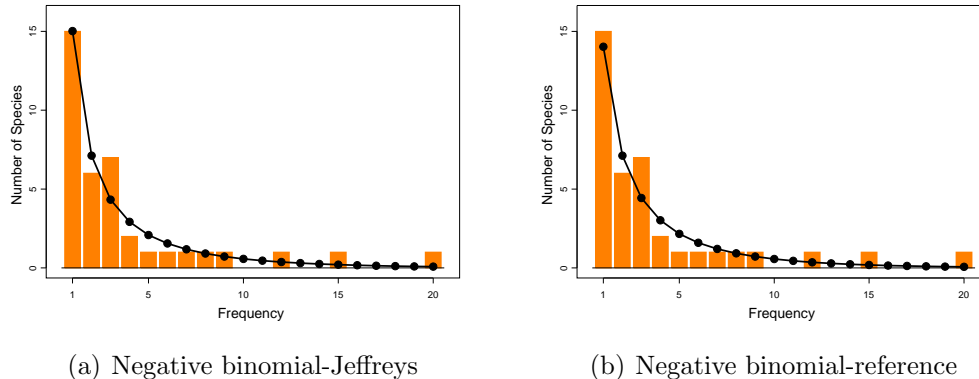


Figure 4.3: Expected frequencies for the negative binomial model.

4.4.2 Lepidoptera

Simulated posteriors contain 500,000 iterations after a 1,000 iteration burn-in period keeping acceptance rates near 30%. Figure 4.4 shows posterior simulations from each of the two priors. The value for τ is set at 45. Diagnostic plots for our samplers are shown in Figures 4.5. For each prior we show trace plots and autocorrelation plots for C . The diagnostic plots for each model show that the sampler has converged.

Bayesian estimates for C are shown in Table 4.2. The median of the posterior sample is considered as a point estimate and a 95% central credible interval is constructed using the 0.025 and 0.975 quantiles of the posterior. Conditional maximum likelihood estimates and log transformed confidence intervals are shown.

We can see that the Bayesian estimates are similar to the maximum likelihood estimates for each model. The point estimate for the Jeffreys prior is larger than

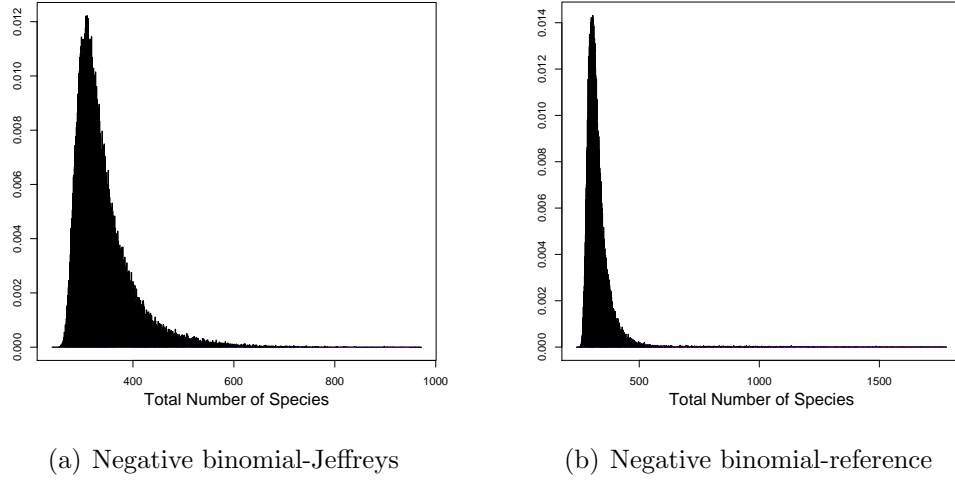
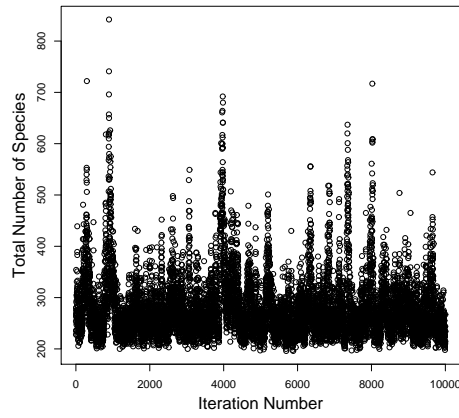


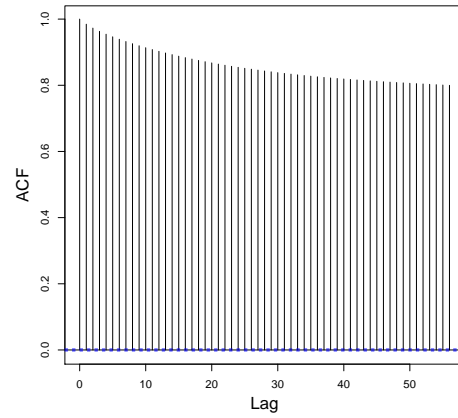
Figure 4.4: Histograms of posterior simulations from $\pi(C|\text{data})$. (a) Negative binomial model with Jeffreys prior. (b) Negative binomial model with reference prior.

Table 4.2: Estimates for the negative binomial model. Bayesian estimates resulting from the Jeffreys and reference priors are shown as well as the maximum likelihood estimate (MLE).

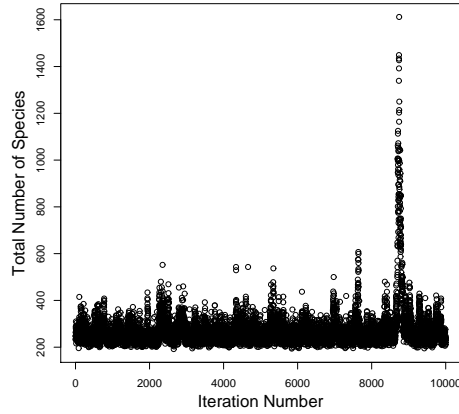
Model	Point estimate	CI
Jeffreys	326	(277,505)
Reference	316	(274,466)
MLE	306.53	(249.01,731.07)



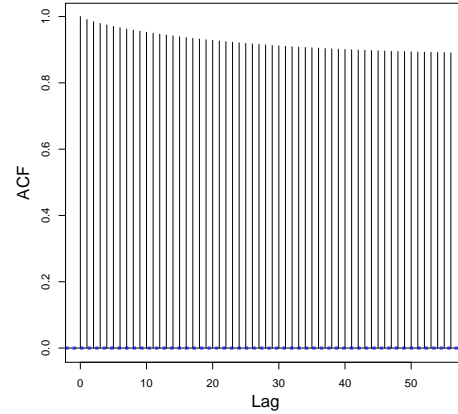
(a) Trace plot for negative binomial-Jeffreys



(b) Autocorrelation plot for negative binomial-Jeffreys



(c) Trace plot for negative binomial-reference



(d) Autocorrelation plot for negative binomial-reference

Figure 4.5: Diagnostic plots for the negative binomial model.

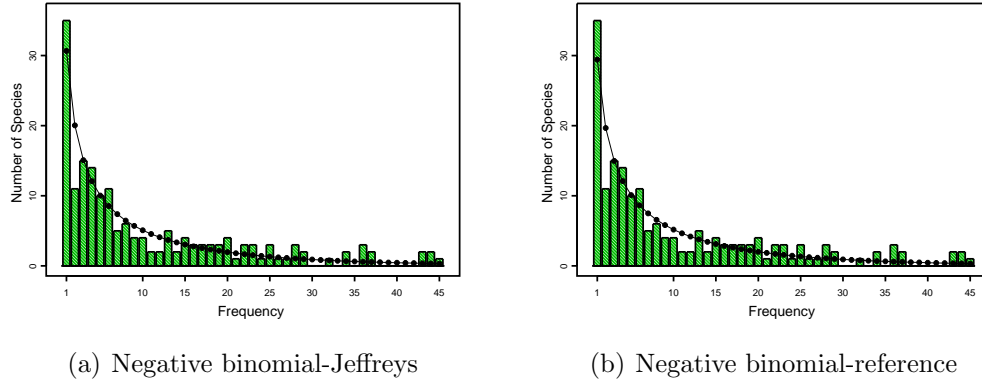


Figure 4.6: Expected frequencies for the negative binomial model.

the point estimate for the reference prior. The confidence interval for the MLE is slightly wider than the Bayesian credible intervals.

We assess the fit of each model using plots of the fitted values. Figure 4.6 shows plots of the raw data with expected values for the frequencies n_1, n_2, \dots, n_τ using the median posterior values from the reference priors. We see that for this data set the negative binomial model's fit is acceptable.

Chapter 5

Three or More Nuisance Parameters

This chapter presents the general derivation of the Jeffreys and reference priors for the species problem. Let the abundance distribution have parameter $\eta = (\eta_1, \dots, \eta_m)$. We will refer to the likelihood in (1.1) and the information matrix in (1.6). Finite mixtures of geometric distributions are used as an extension to the geometric model in Chapter 3. A mixture of two geometrics is implemented in the data analysis section.

5.1 Jeffreys Prior

Derivation of the Jeffreys prior can be achieved by treating the information matrix in (1.6) as a partitioned matrix. The determinant of the information is

$$\det(F(C, \eta))$$

$$\begin{aligned}
&= \left| C \varrho(\eta) \right| \left| \frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} - \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T (C \varrho(\eta))^{-1} \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right) \right| \\
&= C^{m-1} \left| \varrho(\eta) \right| \left| \frac{1 - p_\eta(0)}{p_\eta(0)} - \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T (\varrho(\eta))^{-1} \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right) \right| \tag{5.1}
\end{aligned}$$

Thus, by taking the square root the Jeffreys prior is

$$\pi(C, \eta) \propto C^{\frac{m-1}{2}} \pi(\eta)$$

where $\pi(\eta)$ is determined by (5.1). The Jeffreys prior can be written as a product of two independent priors. For $m \geq 0$ the Jeffreys prior is improper in C . The functions $\pi(C)$ are an increasing sequence in m , and the priors are increasing functions.

The posterior for the model is

$$\begin{aligned}
\pi(C, \eta | \text{data}) &\propto \pi(C, \eta) L(\text{data} | C, \eta) \\
&= C^{\frac{m-1}{2}} \pi(\eta) \frac{C!}{(C-w)!} (p_\eta(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j}
\end{aligned}$$

where $C = w, w+1, w+2, \dots$ and $\eta \in \mathbb{R}$. We need to show

$$\int d\pi(C, \eta | \text{data}) < \infty.$$

We begin with the iterated integral

$$\begin{aligned}
&\int_{\mathbb{R}} \sum_{C \geq w} \pi(C, \eta | x) d\eta \\
&= \int_{\mathbb{R}} \sum_{C \geq w} \pi(C, \eta) \frac{C!}{(C-w)!} (p_\eta(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \\
&= \int_{\mathbb{R}} \pi(\eta) \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \sum_{C \geq w} C^{\frac{m-1}{2}} \frac{C!}{(C-w)!} (p_\eta(0))^{C-w}
\end{aligned}$$

where $\sum_{C \geq w} C^{\frac{m-1}{2}} \frac{C!}{(C-w)!} (p_\eta(0))^{C-w}$ are moments of the negative binomial distribution (without the normalizing constant.) Since all of the moments of the negative binomial exist, the sum is always finite. Closed form solutions can be found for integer values of $\frac{m-1}{2}$ or from $\lceil \frac{m-1}{2} \rceil$ where $\lceil \cdot \rceil$ is the ceiling function and we obtain an upper bound to the marginal posterior,

$$\pi_C(\eta|\text{data}) \propto \pi(\eta) \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \frac{M^{(\lceil \frac{m-1}{2} \rceil)}(0)}{(1 - p_\eta(0))^{w+1}}$$

where $M^{(\lceil \frac{m-1}{2} \rceil)}(0)$ is the $\lceil \frac{m-1}{2} \rceil$ -th derivative of the moment generating function of the negative binomial evaluated at zero, for $m = 2, 3, \dots$. For $m = 0, 1$, let $M^{(\lceil \frac{m-1}{2} \rceil)}(0) = 1$.

5.2 Reference Prior

The reference prior is derived in this section for a model with m nuisance parameters. We will need the inverse of the information in (1.6). Denote this inverse by $S(C, \eta) = F(C, \eta)^{-1}$. Then,

$$S(C, \eta) = \begin{pmatrix} F_{11}^{-1} + F_{11}^{-1} F_{12} E^{-1} F_{21} F_{11}^{-1} & -F_{11}^{-1} F_{12} E^{-1} \\ -E^{-1} F_{21} F_{11}^{-1} & E^{-1} \end{pmatrix}$$

where $E = F_{22} - F_{21} F_{11}^{-1} F_{12}$, and F_{ij} , $i, j = 1, 2$ are the elements of the partitioned information matrix. Now,

$$\begin{aligned} E &= C \varrho(\eta) - \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right) \left(\frac{1}{C} \frac{1 - p_\eta(0)}{p_\eta(0)} \right)^{-1} \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T \\ &= C \left(\varrho(\eta) - \frac{p_\eta(0)}{1 - p_\eta(0)} \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right) \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T \right) \end{aligned}$$

The elements of this matrix are

$$\begin{aligned} S(C, \eta)_{11} &= C \frac{p_\eta(0)}{1 - p_\eta(0)} + \left(C \frac{p_\eta(0)}{1 - p_\eta(0)} \right) \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T E^{-1} \\ &\quad \times \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right) \left(C \frac{p_\eta(0)}{1 - p_\eta(0)} \right) \\ &= Cs(\eta)_{11}, \end{aligned}$$

$$\begin{aligned}
S(C, \eta)_{12} &= - \left(C \frac{p_\eta(0)}{1 - p_\eta(0)} \right) \left(- \frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T E^{-1} \\
&= s(\eta)_{12},
\end{aligned}$$

$$\begin{aligned}
S(C, \eta)_{21} &= -E^{-1} \left(- \frac{\partial}{\partial \eta} \log p_\eta(0) \right) \left(C \frac{p_\eta(0)}{1 - p_\eta(0)} \right) \\
&= s(\eta)_{21},
\end{aligned}$$

and

$$S(C, \eta)_{22} = E^{-1} = \frac{1}{C} s(\eta)_{22}.$$

Thus, the information matrix has the form

$$S(C, \eta) = \begin{pmatrix} Cs(\eta)_{11} & s(\eta)_{12} \\ s(\eta)_{21} & \frac{1}{C}s(\eta)_{22} \end{pmatrix}.$$

The next result follows from proposition 3 in Bernardo and Ramón (1998). If S_j is the $(j \times j)$ upper matrix of $S(C, \eta)$ and $H_j = S_j^{-1}$, then each H_j is a $(j \times j)$ matrix with the form

$$H_j = \begin{pmatrix} \frac{1}{C}h(\eta)_1 & h(\eta)_2 \\ h(\eta)_3 & CH(\eta) \end{pmatrix}$$

where $h(\eta)_1$ is a scalar, $h(\eta)_2$ is $1 \times (j - 1)$, $h(\eta)_3$ is $(j - 1) \times 1$, and $H(\eta)$ is $(j - 1) \times (j - 1)$.

The conditional reference priors are

$$\pi(\eta_m | C, \eta_1, \dots, \eta_{m-1}) \propto \varrho(\eta)_{mm}^{1/2}$$

and

$$\begin{aligned}
& \pi(\eta_k|C, \eta_1, \dots, \eta_{k-1}) \\
& \propto \exp \left[\int \int \log \left(C h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \\
& = \exp \left[\int \int \left(\log C + \log h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \\
& = \exp \left[\log C \int \int \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \\
& \quad \times \exp \left[\int \int \left(\log h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \quad (5.2) \\
& \propto \exp \left[\int \int \left(\log h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right]
\end{aligned}$$

where $\eta_{\mathbf{k}+1} = d\eta_{k+1} \times \dots \times d\eta_m$ if all of the $\pi(\eta_k|C, \eta_1, \dots, \eta_{k-1})$, $k = 1, \dots, m$ are proper. If any of the conditional reference priors are not proper, than a compact approximation is required for the corresponding integrals. In (5.2) we are able to consider the first exponential as a constant with respect to η_k since $\left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} = 1$ if all of these conditional priors are proper or if a compact approximation is used. This means all of the conditional priors for η are functions of η only. In fact, the conditional priors are the reference priors for C i.i.d. replicates from p_η .

The marginal reference prior for C is

$$\begin{aligned}
\pi(C) & \propto \exp \left[\int \dots \int \log \left(C^{-1/2} s(\eta)_{11}^{-1/2} \right) \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
& = \exp \left[\int \dots \int \left(\log C^{-1/2} + \log s(\eta)_{11}^{-1/2} \right) \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
& = \exp \left[\log C^{-1/2} \int \dots \int \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
& \quad \times \exp \left[\int \dots \int \left(\log s(\eta)_{11}^{-1/2} \right) \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
& \propto C^{-1/2}
\end{aligned}$$

where all of the conditional priors are proper or a compact approximation is used for the corresponding integrals.

5.3 A Three Parameter Mixture Model

Let $f(\lambda|\alpha, \theta_1, \theta_2)$ be a mixture of two exponential distributions parameterized as

$$f(\lambda|\alpha, \theta_1, \theta_2) = \alpha \frac{1}{\theta_1} e^{-\lambda/\theta_1} + (1 - \alpha) \frac{1}{\theta_2} e^{-\lambda/\theta_2}$$

with $\theta_1, \theta_2 > 0$, $0 \leq \alpha \leq 1$, and $0 \leq \lambda < \infty$. Then $E[\Lambda] = \alpha\theta_1 + (1 - \alpha)\theta_2$. The two mixed exponential-mixed Poisson distribution is

$$\begin{aligned} p_{\alpha, \theta_1, \theta_2}(x) &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \left(\alpha \frac{1}{\theta_1} e^{-\lambda/\theta_1} + (1 - \alpha) \frac{1}{\theta_2} e^{-\lambda/\theta_2} \right) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \alpha \frac{1}{\theta_1} e^{-\lambda/\theta_1} d\lambda + \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \alpha \frac{1}{\theta_2} e^{-\lambda/\theta_2} d\lambda \\ &= \alpha \frac{1}{1 + \theta_1} \left(\frac{\theta_1}{1 + \theta_1} \right)^x + (1 - \alpha) \frac{1}{1 + \theta_2} \left(\frac{\theta_2}{1 + \theta_2} \right)^x \end{aligned}$$

which is a mixture of two geometric distributions. Using this mixture as the abundance distribution in the species models allows for more flexible modeling of the data.

The likelihood for this model is

$$\begin{aligned} L(\text{data}|C, \theta_1, \theta_2, \alpha) &= \frac{C!}{(C - w)!} \left(\alpha \frac{1}{1 + \theta_1} + (1 - \alpha) \frac{1}{1 + \theta_2} \right)^{(C - w)} \frac{1}{\prod_{j \geq 1} n_j!} \\ &\quad \times \prod_{j \geq 1} \left[\alpha \frac{1}{1 + \theta_1} \left(\frac{\theta_1}{1 + \theta_1} \right)^j \right. \\ &\quad \left. + (1 - \alpha) \frac{1}{1 + \theta_2} \left(\frac{\theta_2}{1 + \theta_2} \right)^j \right]^{n_j}. \end{aligned}$$

5.4 Data Analysis

The mixture of two geometrics model is used to model the abundance distributions for the Framvaren Fjord and Lepidoptera data. This model's abundance

distribution has three parameters. The mixture of two geometric distributions is monotonically decreasing as opposed to the negative binomial. However, the monotonic decrease of the abundance distribution is not much of a restriction since most data sets of this kind have a very large number of singletons (number of species represented in the sample by one individual) compared to the other frequencies.

In order to simplify the form of the prior, we recommend the use of the marginal prior for the number of species and a different objective marginal prior for the nuisance parameters. The mixture of two geometric distributions has parameters where θ_1 and θ_2 are both parameters for the geometric distribution and α is the mixing proportion. In this implementation we chose independent priors for the parameters of the geometric distribution; namely,

$$\pi(\theta_1) \propto \theta_1^{-1/2}(1 + \theta_1)^{-1}$$

and

$$\pi(\theta_2) \propto \theta_2^{-1/2}(1 + \theta_2)^{-1}.$$

Recall that the reference prior for C independent geometric random variables is $\theta^{-1/2}(1 + \theta)^{-1/2}$. The prior we have chosen for θ_1 and θ_2 are similar to the reference prior; however, we do not use the reference prior since the full conditional is improper in some instances of the sampler. The parameter for the mixing proportion is given a uniform prior on $(0, 1)$.

Simulation from the posterior for each model uses a Gibbs sampler. Since the data is i.i.d. when conditional on C , we take advantage of the mixture distribution by using data augmentation techniques for mixed models (Gelman et al. 2004; Tanner & Wong 1987) on the full conditional for the nuisance parameters. The technique is similar to the Expectation-Maximization algorithm for mixture

models. Data augmentation for Bayesian models involves assuming the data is (X, Z) where Z is a vector of indicator variables which indicate which component of the mixture each data point comes from. Z is treated as unknown in the model and is added to the algorithm and sampled along with the other parameters in the model. We sample alternatively from the full conditional distributions, $\pi(\theta_1, \theta_2, \alpha | C, \text{data})$ and $\pi(C | \theta_1, \theta_2, \alpha, \text{data})$.

Our implementation has utilized the Gibbs sampler by sampling from the full conditionals $\pi(C | \eta, \text{data})$ and $\pi(\eta | C, \text{data})$. This is a very useful algorithm since the full conditional for the nuisance parameters is conditional on C , allowing formulation as an i.i.d. problem. This leaves avenues open for techniques such as hierarchical modeling, which could be helpful in facilitating the objective Bayesian approach. The convergence is slow for this algorithm due to the correlation between C and $p_\eta(0)$ (Wang et al. 2007), so many iterations are required.

For the two-mixed geometric model using the Jeffreys prior, $\pi(C) \propto C$, the posterior we are sampling from is

$$\begin{aligned} \pi(C, \theta_1, \theta_2, \alpha | \text{data}) & \propto \pi(C, \theta_1, \theta_2, \alpha) L(\text{data} | C, \theta_1, \theta_2, \alpha) \\ & \propto C \theta_1^{-1/2} (1 + \theta_1)^{-1} \theta_2^{-1/2} (1 + \theta_2)^{-1} \frac{C!}{(C - w)!} \frac{1}{\prod_{j \geq 1} n_j!} \\ & \quad \times \left(\alpha \frac{1}{1 + \theta_1} + (1 - \alpha) \frac{1}{1 + \theta_2} \right)^{(C - w)} \\ & \quad \times \prod_{j \geq 1} \left[\alpha \frac{1}{1 + \theta_1} \left(\frac{\theta_1}{1 + \theta_1} \right)^j + (1 - \alpha) \frac{1}{1 + \theta_2} \left(\frac{\theta_2}{1 + \theta_2} \right)^j \right]^{n_j} \end{aligned}$$

and the full conditionals are

$$\pi(\theta_1, \theta_2, \alpha | C, \text{data})$$

$$\propto \theta_1^{-1/2}(1+\theta_1)^{-1}\theta_2^{-1/2}(1+\theta_2)^{-1} \left(\alpha \frac{1}{1+\theta_1} + (1-\alpha) \frac{1}{1+\theta_2} \right)^{(C-w)} \\ \times \prod_{j \geq 1} \left[\alpha \frac{1}{1+\theta_1} \left(\frac{\theta_1}{1+\theta_1} \right)^j + (1-\alpha) \frac{1}{1+\theta_2} \left(\frac{\theta_2}{1+\theta_2} \right)^j \right]^{n_j}$$

and

$$\pi(C|\theta_1, \theta_2, \alpha, \text{data}) \propto C\theta_1^{-1/2} \frac{C!}{(C-w)!} \left(\alpha \frac{1}{1+\theta_1} + (1-\alpha) \frac{1}{1+\theta_2} \right)^C.$$

The full conditional for the nuisance parameters is not used directly. The likelihood for the model when C is known is an i.i.d. sample from a mixture of two geometric distributions. For augmented data (X, Z) where $X = (x_1, \dots, x_C)$ is the full data (including zero counts since C is known) and $Z = (z_1, \dots, z_C)$ is a vector of indicator variables for the first component in the mixture, the likelihood for the augmented data can be written as

$$L(X, Z|\theta_1, \theta_2, \alpha) = \prod_{i=1}^C \alpha^{z_i} \left[\frac{1}{1+\theta_1} \left(\frac{\theta_1}{1+\theta_1} \right)^{x_i} \right]^{z_i} (1-\alpha)^{1-z_i} \\ \times \left[\frac{1}{1+\theta_2} \left(\frac{\theta_2}{1+\theta_2} \right)^{x_i} \right]^{1-z_i}$$

and the full conditionals are

$$\pi(z_i|\theta_1, \theta_2, \alpha, X) \\ \propto \left[\alpha \frac{1}{1+\theta_1} \left(\frac{\theta_1}{1+\theta_1} \right)^{x_i} \right]^{z_i} \left[(1-\alpha) \frac{1}{1+\theta_2} \left(\frac{\theta_2}{1+\theta_2} \right)^{x_i} \right]^{1-z_i} \\ = \text{Bernoulli} \left(\alpha \frac{1}{1+\theta_1} \left(\frac{\theta_1}{1+\theta_1} \right)^{x_i} \right. \\ \left. / \left[\alpha \frac{1}{1+\theta_1} \left(\frac{\theta_1}{1+\theta_1} \right)^{x_i} (1-\alpha) \frac{1}{1+\theta_2} \left(\frac{\theta_2}{1+\theta_2} \right)^{x_i} \right] \right),$$

$$\pi(\alpha|\theta_1, \theta_2, X, Z) \propto \alpha^{\sum z_i} (1-\alpha)^{C-\sum z_i} \\ = \text{beta} \left(\sum z_i + 1, C - \sum z_i + 1 \right),$$

$$\pi(\theta_1|\theta_2, \alpha, X, Z) \propto \theta_1^{-1/2}(1 + \theta_1)^{-1} \left(\frac{1}{1 + \theta_1} \right)^{\sum z_i} \left(\frac{\theta_1}{1 + \theta_1} \right)^{\sum x_i z_i},$$

and

$$\pi(\theta_2|\theta_1, \alpha, X, Z) \propto \theta_2^{-1/2}(1 + \theta_2)^{-1} \left(\frac{1}{1 + \theta_2} \right)^{C - \sum z_i} \left(\frac{\theta_2}{1 + \theta_2} \right)^{\sum x_i - \sum x_i z_i}.$$

The full conditionals for θ_1 and θ_2 can be reparameterized to obtain

$$\frac{\theta_1}{1 + \theta_1} \sim \text{beta} \left(\sum x_i z_i + 1/2, \sum z_i + 1/2 \right)$$

and

$$\frac{\theta_2}{1 + \theta_2} \sim \text{beta} \left(\sum x_i - \sum x_i z_i + 1/2, C - \sum z_i + 1/2 \right).$$

All of the full conditionals for the nuisance parameters are proper and can be sampled from directly.

5.4.1 Framvaren Fjord

Simulated posteriors contain 1,000,000 iterations after a 1,000 iteration burn-in period keeping acceptance rates near 30%. Figure 5.1 shows posterior simulations from each of the two priors. The value for τ is set at 165. Diagnostic plots for our samplers are shown in Figures 5.2. For each prior we show trace plots and autocorrelation plots for C . The diagnostic plots for each model show that we are confident our MCMC has converged. Also, the autocorrelation plots look much better than for the negative binomial model.

Bayesian estimates for C are shown in Table 5.1. The median of the posterior sample is considered as a point estimate and a 95% central credible interval is constructed using the 0.025 and 0.975 quantiles of the posterior. Conditional maximum likelihood estimates and log transformed confidence intervals are shown.

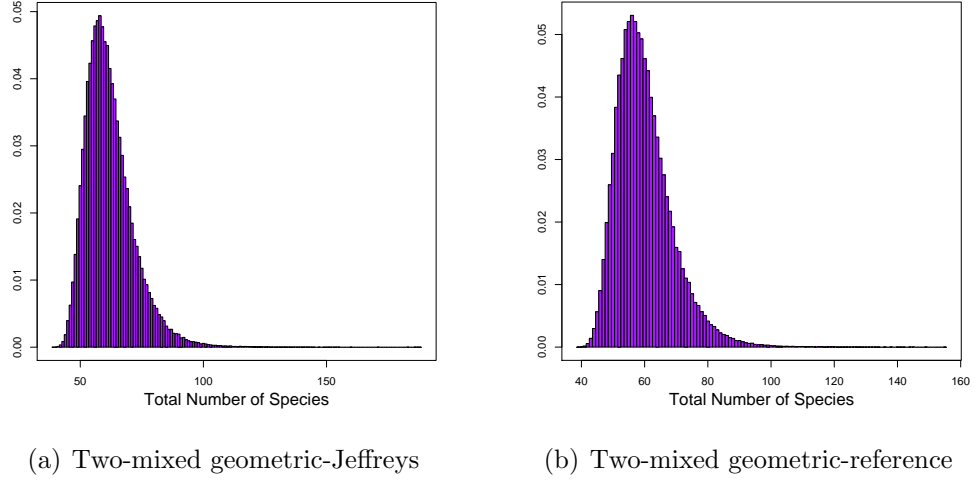
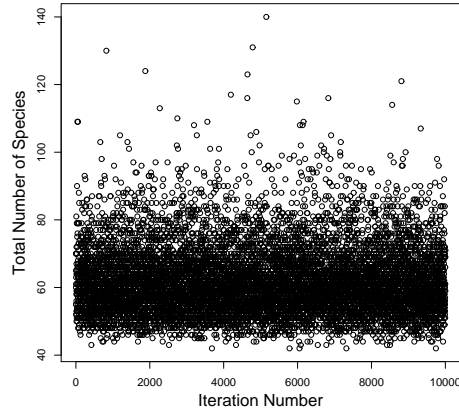


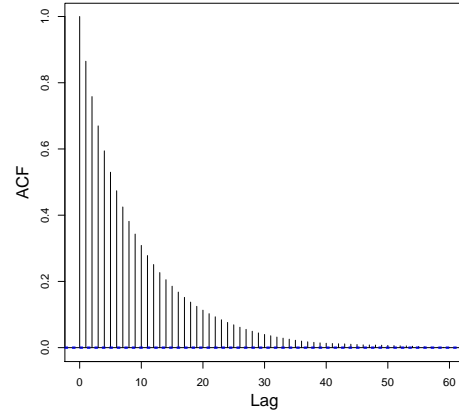
Figure 5.1: Histograms of posterior samples from $\pi(C|\text{data})$. (a) Negative binomial model with Jeffreys prior. (b) Negative binomial model with reference prior.

Table 5.1: Estimates for the two-mixed geometric model. Bayesian estimates resulting from the Jeffreys and reference priors are shown as well as the maximum likelihood estimate (MLE).

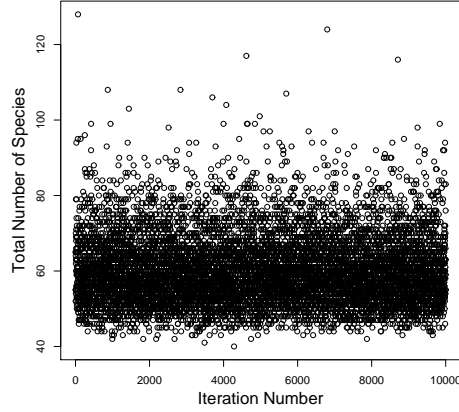
Model	Point estimate	CI
Jeffreys	60	(48,85)
Reference	59	(47,81)
MLE	56.20	(47.24,74.87)



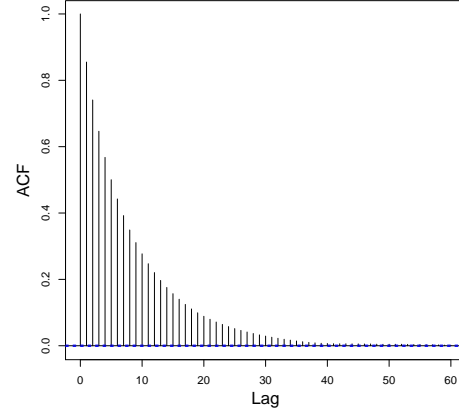
(a) Trace plot for two-mixed geometric-Jeffreys



(b) Autocorrelation plot for two-mixed geometric-Jeffreys



(c) Trace plot for two-mixed geometric-reference



(d) Autocorrelation plot for two-mixed geometric-reference

Figure 5.2: Diagnostic plots for the two-mixed geometric model.

We can see that all of the estimates are all very close. The estimates for the Jeffreys prior are slightly larger than the point estimates for the reference prior.

We asses the fit of each model using plots of the fitted values. Figure 5.3 shows plots of the raw data with expected values for the frequencies n_1, n_2, \dots, n_τ using the median posterior values from the reference priors. We see that for this data set the two-mixed geometric model's fit is acceptable. The fit appears similar to the negative binomial fits in Figure 4.3.

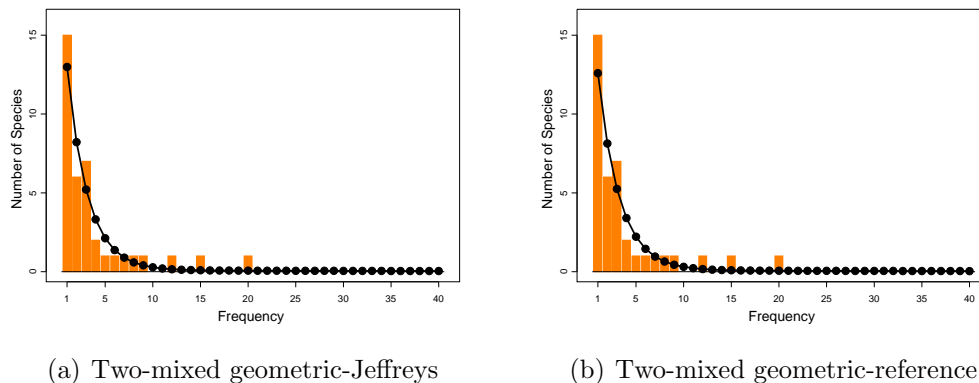


Figure 5.3: Expected frequencies for the two-mixed geometric model.

5.4.2 Lepidoptera

The analysis for this data set with the two-mixed geometric model includes the reference prior only. The sampler using the Jeffreys prior does not converge due to continued sampling of very large values of C . The simulated posterior contains 350,000 iterations after a 1,000 iteration burn-in period keeping acceptance rates near 30%. Figure 5.4 shows the posterior simulation for the reference prior. The value for τ is set at 45. The posterior for the reference prior has a very long tail.

Diagnostic plots for our sampler are shown in Figures 5.5. We show a trace

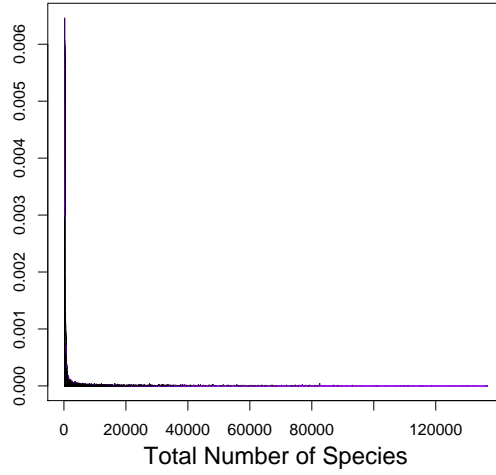


Figure 5.4: Posterior sample from $\pi(C|data)$ using the reference prior.

plot and an autocorrelation plot for C . The diagnostic plots show that the MCMC has converged. Notice the high correlations in the diagnostic plots similar to the other models with the reference prior. The spikes in the trace plot are typical due to the high correlations in the sampler. Care must be taken to ensure an adequate number of iterations.

Bayesian estimates for C are shown in Table 5.2. The median of the posterior sample is considered as a point estimate and a 95% central credible interval is constructed using the 0.025 and 0.975 quantiles of the posterior. Conditional maximum likelihood estimates and log transformed confidence intervals are shown.

The estimates are all different here. The credible interval for the reference prior has a very large upper limit, a consequence from sampling some very large posterior values. The MLE gives a stable, but possibly unreasonable estimate here.

We assess the fit of the model using a plot of the fitted values. Figure 5.6 shows

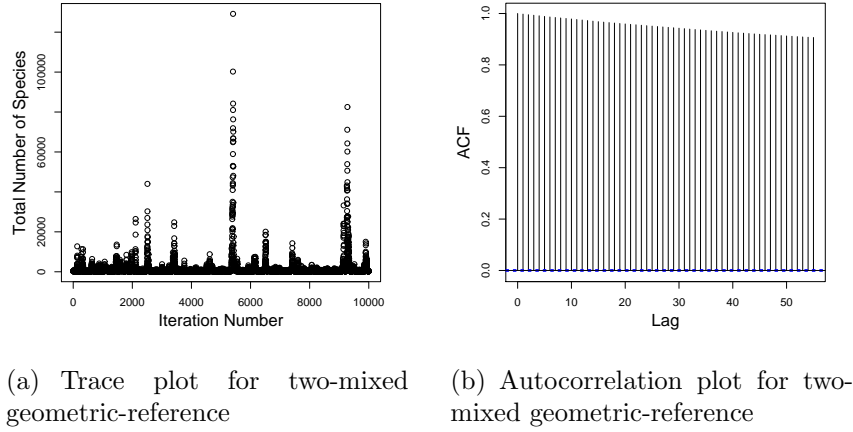


Figure 5.5: Diagnostic plots for the two-mixed geometric model using the reference prior.

Table 5.2: Estimates for the two-mixed geometric model. Bayesian estimates resulting from the Jeffreys and reference priors are shown as well as the maximum likelihood estimate (MLE). NA=did not converge

Model	Point estimate	CI
Jeffreys	NA	NA
Reference	423	(274,7757)
MLE	691.32	(257.22,12071.29)

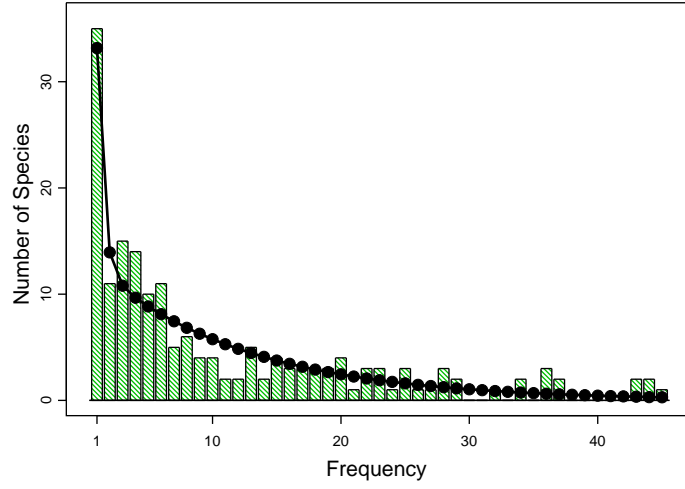


Figure 5.6: Expected frequencies for the two-mixed geometric model using the reference prior.

a plot of the raw data with expected values for the frequencies n_1, n_2, \dots, n_τ using the median posterior values from the reference prior. We see that for this data set the fit is the best we have seen out of all the models.

Chapter 6

Summary and Conclusions

This dissertation has presented a fully Bayesian method using objective priors for the problem of estimating the number of classes in a population. The priors are derived using the methods of Jeffreys (1946) and Bernardo (1979) to generate what we call Jeffreys and reference priors, respectively. These priors are each based on notions of objectiveness which justify their use as objective priors.

The prior assumes the parameters are independent; i.e. $\pi(C, \eta) = \pi(C)\pi(\eta)$, although the method we use to derive the priors does not make this restriction. This serves as justification for using marginal objective priors in practice.

Full joint priors can be derived for models with one nuisance parameter. These examples can be dealt with analytically due to their simplicity. Models which are slightly more complex are necessary when dealing with larger data sets. When taking a parametric approach, we want to have a variety of models at hand. Our suggestion for assigning a prior $\pi(\eta)$, when the nuisance parameter η is a vector, is arbitrary and requires more investigation.

6.1 Comparison of Jeffreys Prior and Reference Prior

The choice between Jeffreys and reference prior can be based on one's belief in their respective notions of objectiveness and their performance in statistical analysis.

The Jeffreys prior is a function of the number of nuisance parameters, m . Although it is inappropriate to interpret an improper prior as a description of belief since it is not a probability density function, improper priors are often interpreted as belief functions. For instance, the prior $\pi(\phi) \propto 1$ where $\phi > 0$ is often viewed as believing all values of ϕ are equally likely and is even named an improper *uniform* prior on ϕ . The case of the Jeffreys prior for the species problem is quite different in that the prior is an increasing function in C . The interpretation of the Jeffreys prior is not obvious. If we interpret the Jeffreys prior as a limit to a proper prior on a bounded parameter space, this means larger values of C are always more likely.

On the other hand, the reference prior is constant across all models. Reference priors use information on the order of importance between the parameters in a problem. The reference prior has been successful over the Jeffreys prior in other multivariate problems (Irony 1997). From the analysis results, it appears the reference prior is favorable to the Jeffreys prior for models with a larger number of nuisance parameters.

From our data analysis results we can see that for problems with multiple nuisance parameters, Bayesian point estimates for the number of species using the Jeffreys prior are larger than estimates under the same model using the reference prior. This reflects the fact that the Jeffreys prior is the dominating function.

6.2 Comments on Bayesian, Frequentist, and Nonparametric Approaches

The approach in this dissertation is parametric. However, the method could be applied to a nonparametric model for species estimation described in Wang and Lindsay (2005), Böhning and Schön (2005), and Norris and Pollock (1998). A Bayesian version of the nonparametric model would be similar to the methods described in this paper with the addition of placing a prior on the number of model parameters. This approach is feasible and is an area of future research.

In the data analysis the maximum likelihood estimates are comparable with estimates from the reference prior, but not with the Jeffreys prior except in the simplest models in Chapter 3.

6.3 Model Selection

We have used plots of fitted values to assess the absolute fit of the model. However, a more quantitative model selection procedure is desired.

We have implemented several models with varying complexity on our example data sets. Determining the best model in the species problem is complicated since model selection and the subsetting problem both pose difficulties. Bayes factors can not be directly used for model selection since they are ill defined when improper priors are used. We instead use deviance to compare models, although this likelihood-based statistic can only be used for a common subset of the data (equal τ).

In order to check the relative fit of our models we have computed the deviance for each model, averaged over values from the posterior sample. Deviance is

Table 6.1: Deviances for Framvaren Fjord data at $\tau = 20$. NB-J=Negative binomial model with Jeffreys prior, NB-R=Negative binomial model with reference prior, MG-J=Two-mixed geometric model with Jeffreys prior, and MG-R=Two-mixed geometric model with reference prior.

Model	Deviance
NB-J	46.158
NB-R	46.284
MG-J	47.952
MG-R	47.224

defined as

$$D(\text{data}, C, \eta) = -2 \log L(\text{data}|C, \eta).$$

We estimate the expected deviance by averaging the deviance over the posterior samples, (C^i, η^i) , $i = 1, \dots, R$ with R being the total number of posterior samples.

The estimate of the deviance is

$$\hat{D}(\text{data}) = \frac{1}{R} \sum_{i=1}^R D(\text{data}, C^i, \eta^i). \quad (6.1)$$

Deviance estimates are provided in Tables 6.1 and 6.2. Lower deviance reflects a better fit, so the two-mixed geometric model fits the data better at $\tau = 165$ than the negative binomial model. At the smaller $\tau = 20$, there is not much difference in the deviance for the two models. Also, the deviances for a particular model-prior combination are virtually equal.

Choosing τ is an important part of the model selection procedure. See Hong et al. (2006) for a frequentist-based model selection routine. For the Bayesian estimates deviance does not work when comparing analyses of *different* subsets. The ideal solution to the subsetting problem would be to use an estimator robust to the species with large abundances.

Table 6.2: Deviances for Framvaren Fjord data at $\tau = 165$. NB-J=Negative binomial model with Jeffreys prior, NB-R=Negative binomial model with reference prior, MG-J=Two-mixed geometric model with Jeffreys prior, and MG-R=Two-mixed geometric model with reference prior.

Model	Deviance
NB-J	73.27935
NB-R	73.99615
MG-J	65.48426
MG-R	65.23959

6.4 Use of the Method in Practice

In order to conduct this estimation using an objective Bayesian approach one can do so by following the general form of the priors in Chapter 5 and applying them to specific models. In cases where the prior is difficult to compute, one can use the marginal prior for the number of species $\pi(C)$ in combination with another objective prior for the nuisance parameter. In this paper, we implement the method for models with one nuisance parameter. Unfortunately, models with even two nuisance parameters, such as the gamma-mixed Poisson model, become difficult to handle and we suggest an alternative objective prior on the nuisance parameter. If proper priors are not used, the posterior must be checked for integrability.

The use of mixture models is an excellent approach to modeling abundance distributions in the species problem. Although the number of components in a mixture model needs to be selected, mixture distributions form a very flexible class of models. We suggest using mixtures of geometric distributions since any monotonically decreasing abundance distribution can be approximated by a mixture of geometric distributions. There is also potential to use separate mixture components to identify subpopulations within a population of interest.

APPENDIX

This appendix contains code for the implementation of the procedures discussed in this dissertation run in the statistical software program R version 2.6.1. with installed packages MASS and HH. The input includes:

- tau: τ
- burn.in: number of burn-in iterations performed
- iterations: number of iterations performed after burn-in
- Metropolis.start.: starting value in metropolis algorithm for the indicated parameter
- Metropolis.stdev.: standard deviation in metropolis algorithm for the indicated parameter
- Metropolis.stdev.C: tuning parameter for Metropolis-Hastings algorithm for C
- filename: location of tab delimited data with first column for frequencies and second column for observed counts
- bars: width of bars in histogram for Cbars; must be an integer
- filename.trace: location for .eps output file
- filename.AC: location for .eps output file
- filename.C: location for .eps output file
- filename.Cbars: location for .eps output file

- filename.analysis: location for .txt output file
- filename.fits: location for .txt output file

The output includes:

- filename.trace: trace plot of posterior samples of C
- filename.AC: autocorrelation plot of posterior samples of C
- filename.C: histogram of posterior samples of C
- filename.Cbars: histogram with specified bar width of posterior samples of C
- filename.analysis: includes the following
 - w: number of observed species in the full data
 - n: number of observed individuals in the full data
 - MLE.est.: unbiased estimate for the indicated parameter
 - w.tau: number of observed species based on τ
 - n.tau: number of observed individuals based on τ
 - acceptance.rate.: acceptance rate for the indicated parameter
 - mode.C: mode of posterior samples of C
 - median.C: median of posterior samples of C
 - mean.C: mean of posterior samples of C
 - LCI.C: 0.025 quantile of posterior samples of C
 - UCI.C: 0.975 quantile of posterior samples of C
 - mean.D: mean of deviance of posterior samples of C

- median.D: median of deviance of posterior samples of C
- DIC: deviance information criteria
- filename.fits: fitted values (expected frequencies) of the abundance distribution using posterior median estimates of the parameters.

6.5 Jeffreys Prior for the Poisson Model

```
###
### R script
### Equal abundance Poisson model
### joint Jeffreys prior
###
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples
iterations<-50000
# starting value for lambda
Metropolis.start.lambda<-1
# stdev for Metropolis algorithm for lambda
Metropolis.stdev.lambda<-1
# location of data
filename<-"C:/Documents and Settings/MyData.txt"

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of analysis file
```

```

filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate NP estimate of n0 and w and n for the full data
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])
NP.est.n0<-w/(1-fullfreqdata[1,2]/n)-w

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics on subsetted data
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])

### Step 3: calculate posterior

## initialize vectors for lambda, n0, and deviance
# specify total number of iterations
iterations<-iterations+burn.in
L<-c(Metropolis.start.lambda,rep(1,iterations-1))
n0<-rep(0,iterations)
D.post<-rep(0,iterations)

```



```

# to track acceptance rate of lambda
a1<-0

for (i in 2:iterations){

  ## sample from p(lambda|x)

  # propose value for lambda
  L.new<-abs(rnorm(1,mean=L[i-1],sd=Metropolis.stdev.lambda))

  # calculate log of acceptance ratio
  logr<-((.5)*(log(exp(L.new)-L.new-1)-log(L.new))+sum(
    freqdata[,2]*(-L.new+freqdata[,1]*log(L.new)))-(w.tau+1)
    *(log(1-exp(-L.new))))-((.5)*(log(exp(L[i-1])-L[i-1]-1)
    -log(L[i-1]))+sum(freqdata[,2]*(-L[i-1]+freqdata[,1]
    *log(L[i-1])))-(w.tau+1)*(log(1-exp(-L[i-1]))))

  # calculate acceptance ratio
  r<-exp(logr)

  # accept or reject propsed value
  if (runif(1)<min(r,1)) {L[i]<-L.new ; a1<-a1+1}
  else L[i]<-L[i-1]

  ## sample from p(C|lambda,x)
  n0[i]<-rnbino(1,size=w.tau+1,prob=1-exp(-L[i]))

  ## calculate deviance from current sample
  D.post[i]<-(-2)*(log(factorial(n0[i]+w.tau))-log(factorial(
    n0[i]))-sum(log(factorial(freqdata[,2])))-L[i]*(n0[i]
    +w.tau)-sum(freqdata[,2]*log(factorial(freqdata[,1]))
    +n.tau*log(L[i]))

}

### Step 4: model diagnostics

## 1) deviance at posterior mean
mean.L<-mean(L[(burn.in+1):iterations])
mean.n0<-mean(n0[(burn.in+1):iterations])

```

```

loglik.post.mean<-log(factorial(mean.n0+w.tau))-log(factorial(
  mean.n0))-sum(log(factorial(freqdata[,2]))) +mean.n0
  *(-mean.L)-mean.L*w.tau+n.tau*log(mean.L)-sum(freqdata[,2]
  *log(factorial(freqdata[,1])))
D.mean<-(-2)*loglik.post.mean

## 2) posterior mean and median deviance
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posterior
median.L<-quantile(L[(burn.in+1):iterations],probs=.5,names=F)
median.n0<-quantile(n0[(burn.in+1):iterations],probs=.5,names=F)

fits<-rep(0,tau)
for (k in 1:tau){
  fits[k]<-(median.n0+w.tau)*dpois(k,median.L)
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,
  count=freqdata[,2]))

### Step 6: results

# needed to use function export.eps
library(HH)
hist.points<-hist(n0[(burn.in+1):iterations],breaks=seq(
  0,max(n0[(burn.in+1):iterations])+1)-0.5)

results<-data.frame(
  filename=filename,

```

```

w=w,
n=n,
NP.est.C=NP.est.n0+w,
tau=tau,
w.tau=w.tau,
n.tau=n.tau,
iterations=iterations,
burn.in=burn.in,
acceptance.rate.lambda=a1/iterations,
mode.C=hist.points$mids[which.max(hist.points$intensities)]+w,
mean.C=mean.n0+w,
median.C=quantile(n0[(burn.in+1):iterations],probs=.5,names=F)
      +w,
LCI.C=quantile(n0[(burn.in+1):iterations],probs=.025,names=F)+w,
UCI.C=quantile(n0[(burn.in+1):iterations],probs=.975,names=F)+w,
mean.D=mean.D,
median.D=median.D,
DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C
# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
N.thin[k]<-n0[k*((iterations-burn.in)/iterations.trace)]+w
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

```

```

# autocorrelation plot for C
my2.acf<-acf(n0[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(n0[(burn.in+1):iterations]+w,breaks=seq(w,max(n0[(burn.in
  +1):iterations])+1+w)-0.5,main="",xlab="",col=5,freq=F,
  ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

```

6.6 Reference Prior for the Poisson Model

```

###
### R script
### Equal abundance Poisson model
### joint reference prior
###
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples
iterations<-50000
# tuning parameter for C. use small value for large var
Metropolis.stdev.N<-20
# starting value for lambda
Metropolis.start.lambda<-1
# stdev for Metropolis algorithm for lambda

```

```

Metropolis.stdev.lambda<-1
# location of data
filename<-"C:/Documents and Settings/MyData.txt"

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate NP estimate of n0
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])
NP.est.n0<-w/(1-fullfreqdata[1,2]/n)-w

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics
w.tau<-sum(freqdata[,2])

```

```

n.tau<-sum(freqdata[,1]*freqdata[,2])

### Step 3: calculate posterior

## initialization
iterations<-iterations+burn.in
N<-rep(0,iterations)
L<-c(Metropolis.start.lambda,rep(1,iterations-1))
# to track acceptance rate of lambda
a1<-0
# to track acceptance rate of C
a2<-0
# starting value based on nonparametric estimate of n0
N[1]<-ceiling(NP.est.n0)+w.tau
D.post<-rep(0,iterations)

for (i in 2:iterations){

  ## sample from p(lambda|x,C)

  # propose value for lambda
  L.new<-abs(rnorm(1,mean=L[i-1],sd=Metropolis.stdev.lambda))

  # calculate log of acceptance ratio
  logr1<-(n.tau-1/2)*log(L.new)-L.new*N[i-1]-(n.tau-1/2)
    *log(L[i-1])+L[i-1]*N[i-1]

  # calculate acceptance ratio
  r1<-exp(logr1)

  # accept or reject propsed value
  if (runif(1)<min(r1,1)) {L[i]<-L.new ; a1<-a1+1}
    else L[i]<-L[i-1]

  ## sample from p(C|lambda,x)

  ## make sure N.new >=w.tau
  repeat {
    N.new<-rnbino(1,mu=N[i-1],size=Metropolis.stdev.N)
    if(N.new>w.tau-1)
      break
  }
}

```

```

    }

## calculate log(N.new!/(N.new-w.tau)!)
N3.new<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.new[j+1]<-log(N.new-j)
}
N2.new<-sum(N3.new)

## calculate log(N[i-1]!/(N[i-1]-w.tau)!)
N3<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3[j+1]<-log(N[i-1]-j)
}
N2<-sum(N3)

# calculate log of acceptance ratio
logr2<-(N2.new-(1/2)*log(N.new)-N.new*L[i])-(N2-(1/2)*log(
  N[i-1])-N[i-1]*L[i])+(log(dnbinom(N[i-1],mu=N.new,size=
  Metropolis.stdev.N)))-(log(dnbinom(N.new,mu=N[i-1],size
  =Metropolis.stdev.N)))

# calculate acceptance ratio
r2<-exp(logr2)

# accept or reject propsed value
if (runif(1)<min(r2,1)) {N[i]<-N.new ; a2<-a2+1}
else N[i]<-N[i-1]

## calculate deviance from current sample

# calculate log(N[i]!/(N[i]-w.tau)!)
N3.curr<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.curr[j+1]<-log(N[i]-j)
}
N2.curr<-sum(N3.curr)

# calculate deviance
D.post[i]<-(-2)*(N2.curr-sum(log(factorial(freqdata[,2])))
  -L[i]*(N[i])-sum(freqdata[,2]*log(factorial(freqdata[,1]
  )))+n.tau*log(L[i]))

```

```

}

### Step 4: model diagnostics

## 1) deviance at posterior mean
mean.L<-mean(L[(burn.in+1):iterations])
mean.N<-mean(N[(burn.in+1):iterations])

  ## calculate log(mean.N!/(mean.N-w.tau)!)
  N3.mean<-rep(0,w.tau)
  for (j in 0:(w.tau-1)){
    N3.mean[j+1]<-log(mean.N-j)
  }
  N2.mean<-sum(N3.mean)

loglik.post.mean<-N2.mean-sum(log(factorial(freqdata[,2])))
  -mean.L*mean.N+n.tau*log(mean.L)-sum(freqdata[,2]*log(
    factorial(freqdata[,1])))
D.mean<-(-2)*loglik.post.mean

## 2) posterior mean and median deviances
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posteriors
median.L<-quantile(L[(burn.in+1):iterations],probs=.5,
  names=F)
median.N<-quantile(N[(burn.in+1):iterations],probs=.5,
  names=F)

fits<-rep(0,tau)

```



```

for (k in 1:tau){
  fits[k]<-(median.N)*dpois(k,median.L)
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
  freqdata[,2]))

### Step 6: results

# this library is needed to use function export.eps
library(HH)
hist.points<-hist(N[(burn.in+1):iterations]+w-w.tau,breaks=
  seq(w,max(N)+w-w.tau+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,
  n=n,
  NP.est.C=NP.est.n0+w,
  tau=tau,
  w.tau=w.tau,
  n.tau=n.tau,
  iterations=iterations,
  burn.in=burn.in,
  acceptance.rate.lambda=a1/iterations,
  acceptance.rate.N=a2/iterations,
  mode.C=hist.points$mids[which.max(hist.points$intensities)],
  mean.C=mean(N[(burn.in+1):iterations])+w-w.tau,
  median.C=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.5,
    names=F),
  LCI.C=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.025,
    names=F),
  UCI.C=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.975,
    names=F),
  mean.D=mean.D,
  median.D=median.D,
  DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

```

```

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C if there are more than 10,000
iterations
# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
N.thin[k]<-N[k*((iterations-burn.in)/iterations.trace)]
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

# autocorrelation plot for C
my2.acf<-acf(N[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1)-0.5,main="",xlab="",
  col='purple',freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

```

6.7 Jeffreys Prior for the Geometric Model

```

###
### R script

```

```

### Exponential-mixed Poisson model
### joint Jeffreys prior
###
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples
iterations<-50000
# starting value for theta
Metropolis.start.theta<-1
# stdev for Metropolis algorithm for theta
Metropolis.stdev.theta<-1
# location of data
filename<-"C:/Documents and Settings/MyData.txt"

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

```

```

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate summary statistics on full data
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics and MLE estimate of n0 and C
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])
R.hat<-(n.tau/w.tau-1)
MLE.est.n0<-w.tau/R.hat
MLE.est.C<-MLE.est.n0+w

### Step 3: calculate posterior

# initialize vectors
# specify total number of iterations
iterations<-iterations+burn.in
R<-c(Metropolis.start.theta,rep(1,iterations-1))
n0<-rep(0,iterations)
D.post<-rep(0,iterations)
a1<-0

for (i in 2:iterations){

  ## sample from p(theta|x)

  # propose value for theta
  R.new<-abs(rnorm(1,mean=R[i-1],sd=Metropolis.stdev.theta))

  # calculate log of acceptance ratio
  logr<-((n.tau-w.tau-0.5)*log(R.new)-n.tau*log(R.new+1))

```

```

      -((n.tau-w.tau-0.5)*log(R[i-1])-n.tau*log(R[i-1]+1))

# calculate acceptance ratio
r<-exp(logr)

# accept or reject propsed value
if (runif(1)<min(r,1)) {R[i]<-R.new ; a1<-a1+1}
  else R[i]<-R[i-1]

## sample from p(C|theta,x)
n0[i]<-rnbino(1,size=w.tau+1,prob=R[i]/(R[i]+1))

## calculate deviance from current sample

# calculate log((n0[i]+w.tau)!/n0[i]!)
N3.curr<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.curr[j+1]<-log(n0[i]+w.tau-j)
}
N2.curr<-sum(N3.curr)

# calculate deviance
D.post[i]<-(-2)*(N2.curr-sum(log(factorial(freqdata[,2]))))
  +n.tau*log(R[i])+(-n0[i]-w.tau-n.tau)*log(1+R[i]))
}

```

Step 4: model diagnostics

```

## 1) deviance at posterior mean
mean.R<-mean(R[(burn.in+1):iterations])
mean.n0<-mean(n0[(burn.in+1):iterations])

## calculate log((mean.n0+w.tau)!/mean.n0!)
N3.mean<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.mean[j+1]<-log(mean.n0+w.tau-j)
}
N2.mean<-sum(N3.mean)

loglik.post.mean<-N2.mean-sum(log(factorial(freqdata[,2])))

```

```

      +n.tau*log(mean.R)-(mean.n0+w.tau+n.tau)*log(1+mean.R)
D.mean<-(-2)*loglik.post.mean

## 2) posterior mean and median deviance
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posterior
median.R<-quantile(R[(burn.in+1):iterations],probs=.5,
  names=F)
median.n0<-quantile(n0[(burn.in+1):iterations],probs=.5,
  names=F)

fits<-rep(0,tau)
for (k in 1:tau){
  fits[k]<-(median.n0+w.tau)*dgeom(k,1/(1+median.R))
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
  freqdata[,2]))

### Step 6: results

# this library is needed to use function export.eps
library(HH)
hist.points<-hist(n0[(burn.in+1):iterations],breaks=seq(
  0,max(n0[(burn.in+1):iterations])+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,

```

```

n=n,
MLE.est.C=MLE.est.C,
tau=tau,
w.tau=w.tau,
n.tau=n.tau,
iterations=iterations,
burn.in=burn.in,
acceptance.rate.theta=a1/iterations,
mode.C=hist.points$mids[which.max(hist.points$intensities)]+w,
mean.C=mean.n0+w,
median.C=quantile(n0,probs=.5,names=F)+w,
LCI.C=quantile(n0,probs=.025,names=F)+w,
UCI.C=quantile(n0,probs=.975,names=F)+w,
mean.D=mean.D,
median.D=median.D,
DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C
# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
N.thin[k]<-n0[k*((iterations-burn.in)/iterations.trace)]+w
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

# autocorrelation plot for C

```

```

my2.acf<-acf(n0[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(n0[(burn.in+1):iterations]+w,breaks=seq(w,max(n0[
  (burn.in+1):iterations])+1+w)-0.5,main="",xlab="",col=5,
  freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

```

6.8 Reference Prior for the Geometric Model

```

###
### R script
### Exponential-mixed Poisson model
### joint reference prior
###
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples
iterations<-50000
# tuning parameter large -> small var
Metropolis.stdev.N<-2
# starting value for theta
Metropolis.start.theta<-1
# stdev for Metropolis algorithm for theta
Metropolis.stdev.theta<-1
# location of data

```



```

filename<-"C:/Documents and Settings/MyData.txt"

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate summary statistics on full data
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics and MLE estimate of n0 and C
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])
R.hat<-(n.tau/w.tau-1)
MLE.est.n0<-w.tau/R.hat

```

```

MLE.est.N<-MLE.est.n0+w

### Step 3: calculate posterior

## initialization
iterations<-iterations+burn.in
N<-rep(0,iterations)
R<-c(Metropolis.start.theta,rep(1,iterations-1))
# to track acceptance rate of theta
a1<-0
# to track acceptance rate of N
a2<-0
# starting value based on MLE of C
N[1]<-ceiling(MLE.est.N)
D.post<-rep(0,iterations)

for (i in 2:iterations){

  ## sample from p(theta|C,x)

  # propose value for theta
  R.new<-abs(rnorm(1,mean=R[i-1],sd=Metropolis.stdev.theta))

  # calculate log of acceptance ratio
  logr1<-(-N[i-1]-n.tau-1/2)*log(1+R.new)+(n.tau-1/2)
    *log(R.new)-(-N[i-1]-n.tau-1/2)*log(1+R[i-1])-(n.tau
    -1/2)*log(R[i-1])

  # calculate acceptance ratio
  r1<-exp(logr1)

  # accept or reject propsed value
  if (runif(1)<min(r1,1)) {R[i]<-R.new ; a1<-a1+1}
    else R[i]<-R[i-1]

  ## sample from p(C|theta,x)

  ## make sure N.new >=w.tau
  repeat {
    N.new<-rnbino(1,mu=N[i-1],size=Metropolis.stdev.N)
    if(N.new>w.tau-1)

```

```

    break
  }

  ## calculate log(N.new!/(N.new-w.tau!))
  N3.new<-rep(0,w.tau)
  for (j in 0:(w.tau-1)){
    N3.new[j+1]<-log(N.new-j)
  }
  N2.new<-sum(N3.new)

  ## calculate log(N[i-1]!/(N[i-1]-w.tau!))
  N3<-rep(0,w.tau)
  for (j in 0:(w.tau-1)){
    N3[j+1]<-log(N[i-1]-j)
  }
  N2<-sum(N3)

  # calculate log of acceptance ratio
  logr2<-(N2.new-(1/2)*log(N.new)-N.new*log(1+R[i]))
    -(N2-(1/2)*log(N[i-1])-N[i-1]*log(1+R[i]))+(log(
      dnbinom(N[i-1],mu=N.new,size=Metropolis.stdev.N))-(
        log(dnbinom(N.new,mu=N[i-1],size=Metropolis.stdev.N)))

  # calculate acceptance ratio
  r2<-exp(logr2)

  # accept or reject propsed value
  if (runif(1)<min(r2,1)) {N[i]<-N.new ; a2<-a2+1}
    else N[i]<-N[i-1]

## calculate deviance from current sample

  # calculate log(N[i]!/(N[i]-w.tau!))
  N3.curr<-rep(0,w.tau)
  for (j in 0:(w.tau-1)){
    N3.curr[j+1]<-log(N[i]-j)
  }
  N2.curr<-sum(N3.curr)

  # calculate deviance
  D.post[i]<-(-2)*(N2.curr-sum(log(factorial(freqdata[,2])))
    +n.tau*log(R[i])+(-N[i]-n.tau)*log(1+R[i]))

```

```

}

### Step 4: model diagnostics

## 1) deviance at posterior mean
mean.R<-mean(R[(burn.in+1):iterations])
mean.N<-mean(N[(burn.in+1):iterations])

  ## calculate log(mean.N!/(mean.N-w.tau)!)
  N3.mean<-rep(0,w.tau)
  for (j in 0:(w.tau-1)){
    N3.mean[j+1]<-log(mean.N-j)
  }
  N2.mean<-sum(N3.mean)

loglik.post.mean<-N2.mean-sum(log(factorial(freqdata[,2])))
  +n.tau*log(mean.R)-(mean.N+n.tau)*log(1+mean.R)
D.mean<-(-2)*loglik.post.mean

## 2) posterior mean and median deviances
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posterior
median.R<-quantile(R[(burn.in+1):iterations],probs=.5,
  names=F)
median.N<-quantile(N[(burn.in+1):iterations],probs=.5,
  names=F)

fits<-rep(0,tau)
for (k in 1:tau){

```

```

    fits[k]<-(median.N)*dexp(k,1/(1+median.R))
  }
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
  freqdata[,2]))

### Step 6: results

# this library is needed to use function export.eps
library(HH)
hist.points<-hist(N[(burn.in+1):iterations]+w-w.tau,
  breaks=seq(w,max(N)+w-w.tau+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,
  n=n,
  MLE.est.C=MLE.est.N,
  tau=tau,
  w.tau=w.tau,
  n.tau=n.tau,
  iterations=iterations,
  burn.in=burn.in,
  acceptance.rate.theta=a1/iterations,
  acceptance.rate.N=a2/iterations,
  mode.C=hist.points$mids[which.max(hist.points$intensities)],
  mean.C=mean(N[(burn.in+1):iterations])+w-w.tau,
  median.C=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.5,
    names=F),
  LCI.C=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.025,
    names=F),
  UCI.C=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.975,
    names=F),
  mean.D=mean.D,
  median.D=median.D,
  DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

```

```

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C if there are more than 10,000
  iterations
# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
N.thin[k]<-N[k*((iterations-burn.in)/iterations.trace)]
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

# autocorrelation plot for C
my2.acf<-acf(N[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1)-0.5,main="",xlab="",
  col='purple',freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

```

6.9 Jeffreys Prior for the Negative Binomial Model

```

###
### R script

```

```

### Negative Binomial (Gamma-mixed Poisson) model
### Jeffreys prior for C,  $\pi(C)=C^{(1/2)}$ 
### independent cauchy(scale=1) priors for T1T2
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples
iterations<-50000
# tuning parameter large -> small var
Metropolis.stdev.N<-6
# stating values for theta1
Metropolis.start.T1<-(-0.8)
# stdev for Metropolis algorithm for theta1
Metropolis.stdev.T1<-0.01
# stating values for theta2
Metropolis.start.T2<-0.8
# stdev for Metropolis algorithm for theta2
Metropolis.stdev.T2<-0.01
# location of data
filename<-"C:/Documents and Settings/MyData.txt"
# width of bars in histogram for Cbars; must be an integer
bars<-3

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of posterior samples of Cbars plot
filename.Cbars<-"C:/MyData.Cbars.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

```

```

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate NP estimate of n0
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])
NP.est.n0<-w/(1-fullfreqdata[1,2]/n)-w

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])

### Step 3: calculate posterior

# this library is needed to use function mvrnorm
library(MASS)

## initialization
iterations<-iterations+burn.in
N<-rep(0,iterations)
T1T2<-matrix(rep(c(Metropolis.start.T1,Metropolis.start.T2),
  each=iterations),ncol=2)
# to track acceptance rate of T1T2

```



```

a1<-0
# to track acceptance rate of N
a2<-0
# starting value based on nonparametric estimate of n0
N[1]<-ceiling(NP.est.n0)+w.tau
# storage for deviance replicates
D.post<-rep(0,iterations)

for (i in 2:iterations){

  ## sample from p(T1T2|N,x)

  ## propose value T1T2 from a bivariate normal dist.;
  make sure T1T2.new > {-1,0}
  repeat {
    T1T2.new<-mvrnorm(1, c(T1T2[i-1,1],T1T2[i-1,2]), matrix(c(
      Metropolis.stdev.T1,0,0,Metropolis.stdev.T2),nrow=2))
    if(T1T2.new[1]>(-1) & T1T2.new[2]>0)
      break
  }

  # calculate log of acceptance ratio
  logr1<-(-1)*log(T1T2.new[1]^2+2*T1T2.new[1]+2)-log(
    1+T1T2.new[2]^2)+n.tau*log(T1T2.new[2])-(N[i-1]
    *(1+T1T2.new[1])+n.tau)*log(1+T1T2.new[1]
    +T1T2.new[2])+N[i-1]*(1+T1T2.new[1])*log(1+
    T1T2.new[1])+sum(freqdata[,2]*lgamma(1+T1T2.new[1]
    +freqdata[,1]))-w.tau*lgamma(1+T1T2.new[1])+log(
    T1T2[i-1,1]^2+2*T1T2[i-1,1]+2)+log(1+T1T2[i-1,2]^2)
    -n.tau*log(T1T2[i-1,2])+(N[i-1]*(1+T1T2[i-1,1])
    +n.tau)*log(1+T1T2[i-1,1]+T1T2[i-1,2])-N[i-1]*(1+
    T1T2[i-1,1])*log(1+T1T2[i-1,1])-sum(freqdata[,2]
    *lgamma(1+T1T2[i-1,1]+freqdata[,1]))+w.tau*lgamma(
    1+T1T2[i-1,1])

  # calculate acceptance ratio
  r1<-exp(logr1)

  # accept or reject propsed value
  if (runif(1)<min(r1,1)) {T1T2[i,]<-T1T2.new ; a1<-a1+1}
  else T1T2[i,]<-T1T2[i-1,]

  ## sample from p(N|A,G,x)

```

```

## make sure N.new >=w.tau
repeat {
N.new<-rnbinom(1,mu=N[i-1],size=Metropolis.stdev.N)
if(N.new>w.tau-1)
  break
}

## calculate log(N.new!/(N.new-w.tau)!)
N3.new<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3.new[j+1]<-log(N.new-j)
}
N2.new<-sum(N3.new)

## calculate log(N[i-1]!/(N[i-1]-w.tau)!)
N3<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3[j+1]<-log(N[i-1]-j)
}
N2<-sum(N3)

# calculate log of acceptance ratio
logr2<-(1/2)*log(N.new)+N2.new+N.new*(1+T1T2[i,1])
  *log((1+T1T2[i,1])/(1+T1T2[i,1]+T1T2[i,2]))+log(
  dnbinom(N[i-1],mu=N.new,size=Metropolis.stdev.N))
  -(1/2)*log(N[i-1])-N2-N[i-1]*(1+T1T2[i,1])*log((1+
  T1T2[i,1])/(1+T1T2[i,1]+T1T2[i,2]))-log(dnbinom(N.new,
  mu=N[i-1],size=Metropolis.stdev.N))

# calculate acceptance ratio
r2<-exp(logr2)

# accept or reject propsed value
if (runif(1)<min(r2,1)) {N[i]<-N.new ; a2<-a2+1}
  else N[i]<-N[i-1]

## calculate deviance from current sample

# calculate log(N[i]!/(N[i]-w.tau)!)
N3.curr<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3.curr[j+1]<-log(N[i]-j)
}

```

```

    }
    N2.curr<-sum(N3.curr)

# calculate deviance
D.post[i]<-(-2)*(N2.curr+n.tau*log(T1T2[i,2])-(N[i]
  *(1+T1T2[i,1])+n.tau)*log(1+T1T2[i,1]+T1T2[i,2])
+N[i]*(1+T1T2[i,1])*log(1+T1T2[i,1])+sum(freqdata[,2]
  *lgamma(1+T1T2[i,1]+freqdata[,1]))-w.tau*lgamma(1+
  T1T2[i,1])-sum(log(factorial(freqdata[,2])))-sum(
  freqdata[,2]*lgamma(freqdata[,1]+1)))

}

### Step 4: model diagnostics

## 1) deviance at posterior mean
mean.T1<-mean(T1T2[(burn.in+1):iterations,1])
mean.T2<-mean(T1T2[(burn.in+1):iterations,2])
mean.N<-mean(N[(burn.in+1):iterations])

## calculate log(mean.N!/(mean.N-w.tau!))
N3.mean<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.mean[j+1]<-log(mean.N-j)
}
N2.mean<-sum(N3.mean)

D.mean<-(-2)*(N2.mean+n.tau*log(mean.T2)-(mean.N*(1+mean.T1)
  +n.tau)*log(1+mean.T1+mean.T2)+mean.N*(1+mean.T1)*log(1+
  mean.T1)+sum(freqdata[,2]*lgamma(1+mean.T1+freqdata[,1]))
-w.tau*lgamma(1+mean.T1)-sum(log(factorial(freqdata[,2])))-
-sum(freqdata[,2]*lgamma(freqdata[,1]+1)))

## 2) posterior mean and median deviances
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

```

```
## 4) Deviance information criterion
DIC<-2*mean.D-D.mean
```

```
### Step 5: fitted values based on medians of the marginal
posterior
median.T1<-quantile(T1T2[(burn.in+1):iterations,1],probs=.5,
  names=F)
median.T2<-quantile(T1T2[(burn.in+1):iterations,2],probs=.5,
  names=F)
median.N<-quantile(N[(burn.in+1):iterations],probs=.5,
  names=F)

fits<-rep(0,tau)
for (k in 1:tau){
  fits[k]<-(median.N)*dnbinom(k,size=median.T1+1,prob=
    (median.T1+1)/(median.T1+median.T2+1))
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
  freqdata[,2]))
```

```
### Step 6: results
```

```
# this library is needed to use function export.eps
library(HH)
hist.points<-hist(N[(burn.in+1):iterations]+w-w.tau,breaks=
  seq(w,max(N)+w-w.tau+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,
  n=n,
  NP.est.C=NP.est.n0+w,
  tau=tau,
  w.tau=w.tau,
  n.tau=n.tau,
  iterations=iterations,
  burn.in=burn.in,
  acceptance.rate.T1T2=a1/iterations,
  acceptance.rate.N=a2/iterations,
```

```

mode.N=hist.points$mids[which.max(hist.points$intensities)],
mean.N=mean(N[(burn.in+1):iterations])+w-w.tau,
median.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.5,
  names=F),
LCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.025,
  names=F),
UCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.975,
  names=F),
mean.D=mean.D,
median.D=median.D,
DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C if there are more than 10,000
  iterations
# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
  N.thin[k]<-N[k*((iterations-burn.in)/iterations.trace)]
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

# autocorrelation plot for C
my2.acf<-acf(N[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)

```

```

export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1)-0.5,main="",xlab="",
  col='purple',freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

# a histogram with # bars for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1,length=(max(N[
  (burn.in+1):iterations])-w.tau+1)/bars+1)-0.5,main="",
  xlab="",col='purple',freq=F,ylab="",xlim=c(w,quantile(N[
  (burn.in+1):iterations]+w-w.tau,probs=.975,names=F)))
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.Cbars)

```

6.10 Reference Prior for the Negative Binomial Model

```

###
### R script
### Negative Binomial (Gamma-mixed Poisson) model
### reference prior for C
### independent cauchy(scale=1) priors for T1T2
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples

```

```

iterations<-50000
# tuning parameter large -> small var
Metropolis.stdev.N<-6
# stating values for theta1
Metropolis.start.T1<-(-0.8)
# stdev for Metropolis algorithm for theta1
Metropolis.stdev.T1<-0.01
# stating values for theta2
Metropolis.start.T2<-0.8
# stdev for Metropolis algorithm for theta2
Metropolis.stdev.T2<-0.01
# location of data
filename<-"C:/Documents and Settings/MyData.txt"
# width of bars in histogram for Cbars; must be an integer
bars<-3

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of posterior samples of Cbars plot
filename.Cbars<-"C:/MyData.Cbars.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];

```

```

        break}
    }
}

# calculate NP estimate of n0
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])
NP.est.n0<-w/(1-fullfreqdata[1,2]/n)-w

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])

### Step 3: calculate posterior

# this library is needed to use function mvrnorm
library(MASS)

## initialization
iterations<-iterations+burn.in
N<-rep(0,iterations)
T1T2<-matrix(rep(c(Metropolis.start.T1,Metropolis.start.T2),
  each=iterations),ncol=2)
# to track acceptance rate of T1T2
a1<-0
# to track acceptance rate of N
a2<-0
# starting value based on nonparametric estimate of n0
N[1]<-ceiling(NP.est.n0)+w.tau
# storage for deviance replicates
D.post<-rep(0,iterations)

for (i in 2:iterations){

  ## sample from p(T1T2|N,x)

  ## propose value T1T2 from a bivariate normal dist.;
  make sure T1T2.new > {-1,0}

```



```

repeat {
  T1T2.new<-mvrnorm(1, c(T1T2[i-1,1],T1T2[i-1,2]),
    matrix(c(Metropolis.stdev.T1,0,0,Metropolis.stdev.T2),
      nrow=2))
  if(T1T2.new[1]>(-1) & T1T2.new[2]>0)
    break
}

# calculate log of acceptance ratio
logr1<-(-1)*log(T1T2.new[1]^2+2*T1T2.new[1]+2)-log(1+
  T1T2.new[2]^2)+n.tau*log(T1T2.new[2])-(N[i-1]*(1+
  T1T2.new[1])+n.tau)*log(1+T1T2.new[1]+T1T2.new[2])
+N[i-1]*(1+T1T2.new[1])*log(1+T1T2.new[1])+sum(
  freqdata[,2]*lgamma(1+T1T2.new[1]+freqdata[,1]))
-w.tau*lgamma(1+T1T2.new[1])+log(T1T2[i-1,1]^2+2
*T1T2[i-1,1]+2)+log(1+T1T2[i-1,2]^2)-n.tau*log(
T1T2[i-1,2])+(N[i-1]*(1+T1T2[i-1,1])+n.tau)*log(1+
T1T2[i-1,1]+T1T2[i-1,2])-N[i-1]*(1+T1T2[i-1,1])
*log(1+T1T2[i-1,1])-sum(freqdata[,2]*lgamma(1+
T1T2[i-1,1]+freqdata[,1]))+w.tau*lgamma(1+
T1T2[i-1,1])

# calculate acceptance ratio
r1<-exp(logr1)

# accept or reject propsed value
if (runif(1)<min(r1,1)) {T1T2[i,<T1T2.new ; a1<-a1+1}
  else T1T2[i,<-T1T2[i-1,<

## sample from p(N|A,G,x)

## make sure N.new >=w.tau
repeat {
  N.new<-rnbino(1,mu=N[i-1],size=Metropolis.stdev.N)
  if(N.new>w.tau-1)
    break
}

## calculate log(N.new!/(N.new-w.tau)!)
N3.new<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.new[j+1]<-log(N.new-j)
}

```

```

N2.new<-sum(N3.new)

## calculate log(N[i-1]/(N[i-1]-w.tau)!)
N3<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3[j+1]<-log(N[i-1]-j)
}
N2<-sum(N3)

# calculate log of acceptance ratio
logr2<-(-1/2)*log(N.new)+N2.new+N.new*(1+T1T2[i,1])
  *log((1+T1T2[i,1])/(1+T1T2[i,1]+T1T2[i,2]))+log(
  dnbinom(N[i-1],mu=N.new,size=Metropolis.stdev.N))+
  (1/2)*log(N[i-1])-N2-N[i-1]*(1+T1T2[i,1])*log((1+
  T1T2[i,1])/(1+T1T2[i,1]+T1T2[i,2]))-log(dnbinom(
  N.new,mu=N[i-1],size=Metropolis.stdev.N))

# calculate acceptance ratio
r2<-exp(logr2)

# accept or reject propsed value
if (runif(1)<min(r2,1)) {N[i]<-N.new ; a2<-a2+1}
  else N[i]<-N[i-1]

## calculate deviance from current sample

# calculate log(N[i]/(N[i]-w.tau)!)
N3.curr<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3.curr[j+1]<-log(N[i]-j)
}
N2.curr<-sum(N3.curr)

# calculate deviance
D.post[i]<-(-2)*(N2.curr+n.tau*log(T1T2[i,2])-(N[i]*(1+
  T1T2[i,1])+n.tau)*log(1+T1T2[i,1]+T1T2[i,2])+N[i]
  *(1+T1T2[i,1])*log(1+T1T2[i,1])+sum(freqdata[,2]
  *lgamma(1+T1T2[i,1]+freqdata[,1]))-w.tau*lgamma(1+
  T1T2[i,1])-sum(log(factorial(freqdata[,2])))-sum(
  freqdata[,2]*lgamma(freqdata[,1]+1)))
}

```

```

### Step 4: model diagnostics

## 1) deviance at posterior mean
mean.T1<-mean(T1T2[(burn.in+1):iterations,1])
mean.T2<-mean(T1T2[(burn.in+1):iterations,2])
mean.N<-mean(N[(burn.in+1):iterations])

      ## calculate log(mean.N!/(mean.N-w.tau)!)
      N3.mean<-rep(0,w.tau)
      for (j in 0:(w.tau-1)){
        N3.mean[j+1]<-log(mean.N-j)
      }
      N2.mean<-sum(N3.mean)

D.mean<-(-2)*(N2.mean+n.tau*log(mean.T2)-(mean.N*(1+mean.T1)
  +n.tau)*log(1+mean.T1+mean.T2)+mean.N*(1+mean.T1)*log(1+
  mean.T1)+sum(freqdata[,2]*lgamma(1+mean.T1+freqdata[,1]))
  -w.tau*lgamma(1+mean.T1)-sum(log(factorial(freqdata[,2])))
  -sum(freqdata[,2]*lgamma(freqdata[,1]+1)))

## 2) posterior mean and median deviances
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posterior
median.T1<-quantile(T1T2[(burn.in+1):iterations,1],probs=.5,
  names=F)
median.T2<-quantile(T1T2[(burn.in+1):iterations,2],probs=.5,
  names=F)
median.N<-quantile(N[(burn.in+1):iterations],probs=.5,
  names=F)

```

```

fits<-rep(0,tau)
for (k in 1:tau){
  fits[k]<-(median.N)*dnbinom(k,size=median.T1+1,prob=
    (median.T1+1)/(median.T1+median.T2+1))
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
  freqdata[,2]))

```

Step 6: results

```

# this library is needed to use function export.eps
library(HH)
hist.points<-hist(N[(burn.in+1):iterations]+w-w.tau,breaks=
  seq(w,max(N)+w-w.tau+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,
  n=n,
  NP.est.C=NP.est.n0+w,
  tau=tau,
  w.tau=w.tau,
  n.tau=n.tau,
  iterations=iterations,
  burn.in=burn.in,
  acceptance.rate.T1T2=a1/iterations,
  acceptance.rate.N=a2/iterations,
  mode.N=hist.points$mids[which.max(hist.points$intensities)],
  mean.N=mean(N[(burn.in+1):iterations])+w-w.tau,
  median.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.5,
    names=F),
  LCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.025,
    names=F),
  UCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.975,
    names=F),
  mean.D=mean.D,
  median.D=median.D,
  DIC
)

```

```

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

  # first thin values of C if there are more than 10,000
  iterations
  # must be a divisor of (iterations-burn.in)
  iterations.trace<-min(10000,iterations-burn.in)
  N.thin<-rep(0,iterations.trace)
  for (k in 1:iterations.trace){
    N.thin[k]<-N[k*((iterations-burn.in)/iterations.trace)]
  }

  # make trace plot
  plot(1:iterations.trace,N.thin,xlab="",ylab="")
  mtext("Iteration Number",1,line=2.5,cex=1.5)
  mtext("Total Number of Species",2,line=2.5,cex=1.5)
  export.eps(filename.trace)

# autocorrelation plot for C
my2.acf<-acf(N[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1)-0.5,main="",xlab="",
  col='purple',freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

# a histogram with # bars for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1,length=(max(N[

```

```

(burn.in+1):iterations])-(w.tau+1)/(bars+1)-0.5,main="",
xlab="",col='purple',freq=F,ylab="",xlim=c(w,quantile(N[
(burn.in+1):iterations]+w-w.tau,probs=.975,names=F)))
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.Cbars)

```

6.11 Jeffreys Prior for the Two-mixed Geometric Model

```

###
### R script
### Mixture of 2 Geometrics (parameters N, T1, T2, A=alpha)
### Jeffreys prior for C,  $\pi(C)=C$ 
### uniform prior for alpha
###  $\pi(T)=T^{-1/2}*(1+T)^{-1}$ 
###
###

### Step 1: program specifications

# maximum frequency of the data to be analyzed
tau<-10
# number of burn in iterations
burn.in<-1000
# number of posterior samples
iterations<-50000
# tuning parameter for C, controls var of jumping dist
Metropolis.stdev.N<-4
# starting value, MLE of T1
Metropolis.start.T1<-0.5
# stdev for proposal dist for T1
Metropolis.stdev.T1<-0.6
# starting value, MLE of T2
Metropolis.start.T2<-7
# stdev for proposal dist for T2
Metropolis.stdev.T2<-0.6
# location of data

```

```

filename<-"C:/Documents and Settings/MyData.txt"
# width of bars in histogram for Cbars; must be an integer
bars<-3

## names of output files
# location of trace plot
filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate NP estimate of n0
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])
NP.est.n0<-w/(1-fullfreqdata[1,2]/n)-w

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])

```

```

### Step 3: calculate posterior

## initialization
iterations<-iterations+burn.in
A<-rep(0,iterations)
T1<-rep(0,iterations)
T2<-rep(0,iterations)
N<-rep(0,iterations)
# to track acceptance rate of N
a1<-0
# starting value, nonparametric estimate of n0
N[1]<-ceiling(NP.est.n0)+w.tau
# starting value, MLE of T1
T1[1]<-Metropolis.start.T1
# starting value, MLE of T2
T2[1]<-Metropolis.start.T2
A[1]<-0.5
# storage for deviance replicates
D.post<-rep(0,iterations)

for (i in 2:iterations){

  ## sample from  $p(Z|A,T1,T2,X,N)$ 

  ## create a new vector of length  $N[i-1]$ 
  Z<-rep(0,length=N[i-1])

  ## create a full data vector
  X<-c(rep(0,N[i-1]-w.tau),rep(freqdata[,1],times=
    freqdata[,2]))

  ## sample random bernoulli with appropriate success
  prob for each  $Z[k]$ ; do not allow for Z all zeros
  or ones
  for (k in 1:N[i-1]){
    Z[k]<-rbinom(1,1,prob=A[i-1]*(1/(1+T1[i-1]))
      *(T1[i-1]/(1+T1[i-1]))^X[k]/((A[i-1]*(1/(1
      +T1[i-1]))*(T1[i-1]/(1+T1[i-1]))^X[k])+((1-A[i-1])
      *(1/(1+T2[i-1]))*(T2[i-1]/(1+T2[i-1]))^X[k])))
  }
}

```



```

## sample from  $p(A|Z, T1, T2, X, N)$ 

## sample from beta dist
A[i]<-rbeta(1,shape1=sum(Z)+1,shape2=N[i-1]-sum(Z)+1)

## sample from  $p(T1|A, Z, T2, X, N)$  and  $p(T2|A, Z, T1, X, N)$ 

repeat{
## sample  $T1/(1+T1)$  and  $T2/(1+T2)$  from beta dists
T1.trans<-rbeta(1,shape1=sum(X*Z)+0.5,shape2=sum(Z)+0.5)
T2.trans<-rbeta(1,shape1=sum(X)-sum(X*Z)+0.5,shape2=N[i-1]
  -sum(Z)+0.5)

## back transform to T1 and T2
T1[i]<-T1.trans/(1-T1.trans)
T2[i]<-T2.trans/(1-T2.trans)

## keep  $T1 < T2$ 
if(T1[i]<T2[i])
  break
}

## sample from  $p(N|A, T1, T2, Z, X)$ 

## make sure  $N.new \geq w.tau$ 
repeat {
N.new<-rnbino(1,mu=N[i-1],size=Metropolis.stdev.N)
if(N.new>w.tau-1)
  break
}

## calculate  $\log(N.new!/(N.new-w.tau)!)$ 
N3.new<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3.new[j+1]<-log(N.new-j)
}
N2.new<-sum(N3.new)

## calculate  $\log(N[i-1]!/(N[i-1]-w.tau)!)$ 
N3<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3[j+1]<-log(N[i-1]-j)
}

```

```

N2<-sum(N3)

## calculate log of acceptance ratio
logr1<-log(N.new)+N2.new+N.new*log(A[i]*(1/(1+T1[i]))
+(1-A[i])*(1/(1+T2[i]))) -log(N[i-1])-N2-N[i-1]*log(
A[i]*(1/(1+T1[i]))+(1-A[i])*(1/(1+T2[i]))) +log(
dnbinom(N[i-1],mu=N.new,size=Metropolis.stdev.N))
-log(dnbinom(N.new,mu=N[i-1],size=
Metropolis.stdev.N))

## calculate acceptance ratio
r1<-exp(logr1)

## accept or reject the proposed value
if (runif(1)<min(r1,1)) {N[i]<-N.new ; a1<-a1+1}
else N[i]<-N[i-1]

## calculate deviance from current sample

## calculate log(N[i]/(N[i]-w.tau)!)
N3.curr<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
N3.curr[j+1]<-log(N[i]-j)
}
N2.curr<-sum(N3.curr)

# calculate deviance
D.post[i]<-(-2)*(N2.curr+(N[i]-w.tau)*log(A[i]*(1/(1+
T1[i]))+(1-A[i])*(1/(1+T2[i]))) +sum(freqdata[,2]
*log(A[i]*(1/(1+T1[i]))*(T1[i]/(1+T1[i]))^freqdata[,1]
+(1-A[i])*(1/(1+T2[i]))*(T2[i]/(1+T2[i]))^freqdata[,1]))
-sum(log(factorial(freqdata[,2]))))

}

### Step 4: model diagnostics

## 1) deviance at posterior mean
mean.A<-mean(A[(burn.in+1):iterations])
mean.T1<-mean(T1[(burn.in+1):iterations])
mean.T2<-mean(T2[(burn.in+1):iterations])

```

```

mean.N<-mean(N[(burn.in+1):iterations])

  ## calculate log(mean.N!/(mean.N-w.tau!))
  N3.mean<-rep(0,w.tau)
  for (j in 0:(w.tau-1)){
    N3.mean[j+1]<-log(mean.N-j)
  }
  N2.mean<-sum(N3.mean)

D.mean<-(-2)*(N2.curr+(mean.N-w.tau)*log(mean.A*(1/(1+mean.T1))
+(1-mean.A)*(1/(1+mean.T2))))+sum(freqdata[,2]*log(mean.A
*(1/(1+mean.T1))*(mean.T1/(1+mean.T1))^freqdata[,1]+(1
-mean.A)*(1/(1+mean.T2))*(mean.T2/(1+mean.T2))
^freqdata[,1]))-sum(log(factorial(freqdata[,2]))))

## 2) posterior mean and median deviances
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
  names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posterior
median.A<-quantile(A[(burn.in+1):iterations],probs=.5,
  names=F)
median.T1<-quantile(T1[(burn.in+1):iterations],probs=.5,
  names=F)
median.T2<-quantile(T2[(burn.in+1):iterations],probs=.5,
  names=F)
median.N<-quantile(N[(burn.in+1):iterations],probs=.5,
  names=F)

fits<-rep(0,tau)
for (k in 1:tau){
  fits[k]<-(median.N)*(median.A*dgeom(k,prob=1/(1
+median.T1)))+(1-median.A)*dgeom(k,prob=1/(1

```

```

        +median.T2)))
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
  freqdata[,2]))

### Step 6: results

# this library is needed to use function export.eps
library(HH)
hist.points<-hist(N[(burn.in+1):iterations]+w-w.tau,breaks=
  seq(w,max(N)+w-w.tau+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,
  n=n,
  NP.est.C=NP.est.n0+w,
  tau=tau,
  w.tau=w.tau,
  n.tau=n.tau,
  iterations=iterations,
  burn.in=burn.in,
  acceptance.rate.N=a1/iterations,
  mode.N=hist.points$mids[which.max(hist.points$intensities)],
  mean.N=mean(N[(burn.in+1):iterations])+w-w.tau,
  median.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.5,
    names=F),
  LCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.025,
    names=F),
  UCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.975,
    names=F),
  mean.D=mean.D,
  median.D=median.D,
  DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

# output results and fitted values

```

```

write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C if there are more than 10,000
iterations
# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
N.thin[k]<-N[k*((iterations-burn.in)/iterations.trace)]
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

# autocorrelation plot for C
my2.acf<-acf(N[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
mtext("Lag",1,line=2.5,cex=1.5)
mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
(burn.in+1):iterations])+w-w.tau+1)-0.5,main="",xlab="",
col='purple',freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

# a histogram with # bars for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
(burn.in+1):iterations])+w-w.tau+1,length=(max(N[
(burn.in+1):iterations])-w.tau+1)/bars+1)-0.5,main="",
xlab="",col='purple',freq=F,ylab="",xlim=c(w,quantile(N[
(burn.in+1):iterations]+w-w.tau,probs=.975,names=F)))
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()

```

```
export.eps(filename.Cbars)
```

6.12 Reference Prior for the Two-mixed Geometric Model

```
###  
### R script  
### Mixture of 2 Geometrics (parameters N, T1, T2, A=alpha)  
### reference prior for C  
### uniform prior for alpha  
###  $\pi(T)=T^{(-1/2)}*(1+T)^{(-1)}$   
###  
###  
  
### Step 1: program specifications  
  
# maximum frequency of the data to be analyzed  
tau<-10  
# number of burn in iterations  
burn.in<-1000  
# number of posterior samples  
iterations<-50000  
# tuning parameter for C, controls var of jumping dist  
Metropolis.stdev.N<-4  
# starting value, MLE of T1  
Metropolis.start.T1<-0.5  
# stdev for proposal dist for T1  
Metropolis.stdev.T1<-0.6  
# starting value, MLE of T2  
Metropolis.start.T2<-7  
# stdev for proposal dist for T2  
Metropolis.stdev.T2<-0.6  
# location of data  
filename<-"C:/Documents and Settings/MyData.txt"  
# width of bars in histogram for Cbars; must be an integer  
bars<-3  
  
## names of output files  
# location of trace plot
```

```

filename.trace<-"C:/MyData.trace.eps"
# location of autocorrelation plot
filename.AC<-"C:/MyData.AC.eps"
# location of posterior samples of C plot
filename.C<-"C:/MyData.C.eps"
# location of analysis file
filename.analysis<-"C:/MyData.analysis.txt"
# location of analysis file
filename.fits<-"C:/MyData.fits.txt"

### Step 2: format data

entrydata<-read.table(filename,col.names=c("frequency","count"))
rowdim<-dim(entrydata)[1]
fmaxdata<-entrydata[rowdim,1]
fullfreqdata<-cbind(seq(1,fmaxdata),rep(0,fmaxdata))

# make sure freqdata is frequency data with zeros
for (k in 1:fmaxdata){
  for (m in 1:rowdim){
    if (entrydata[m,1]==k) {fullfreqdata[k,2]<-entrydata[m,2];
      break}
  }
}

# calculate NP estimate of n0
w<-sum(fullfreqdata[,2])
n<-sum(fullfreqdata[,1]*fullfreqdata[,2])
NP.est.n0<-w/(1-fullfreqdata[1,2]/n)-w

# subset data below tau
freqdata<-fullfreqdata[1:tau,]

# calculate summary statistics
w.tau<-sum(freqdata[,2])
n.tau<-sum(freqdata[,1]*freqdata[,2])

### Step 3: calculate posterior

## initialization

```

```

iterations<-iterations+burn.in
A<-rep(0,iterations)
T1<-rep(0,iterations)
T2<-rep(0,iterations)
N<-rep(0,iterations)
# to track acceptance rate of N
a1<-0
# starting value, nonparametric estimate of n0
N[1]<-ceiling(NP.est.n0)+w.tau
# starting value, MLE of T1
T1[1]<-Metropolis.start.T1
# starting value, MLE of T2
T2[1]<-Metropolis.start.T2
A[1]<-0.5
# storage for deviance replicates
D.post<-rep(0,iterations)

for (i in 2:iterations){

  ## sample from p(Z|A,T1,T2,X,N)

  ## create a new vector of length N[i-1]
  Z<-rep(0,length=N[i-1])

  ## create a full data vector
  X<-c(rep(0,N[i-1]-w.tau),rep(freqdata[,1],times=
    freqdata[,2]))

  ## sample random bernoulli with appropriate success
  prob for each Z[k]; do not allow for Z all zeros
  or ones
  for (k in 1:N[i-1]){
    Z[k]<-rbinom(1,1,prob=A[i-1]*(1/(1+T1[i-1]))
      *(T1[i-1]/(1+T1[i-1]))^X[k]/((A[i-1]*(1/(1+
      T1[i-1]))*(T1[i-1]/(1+T1[i-1]))^X[k])+((1-A[i-1])
      *(1/(1+T2[i-1]))*(T2[i-1]/(1+T2[i-1]))^X[k])))
  }

  ## sample from p(A|Z,T1,T2,X,N)

  ## sample from beta dist
  A[i]<-rbeta(1,shape1=sum(Z)+1,shape2=N[i-1]-sum(Z)+1)

```



```

## sample from  $p(T1|A,Z,T2,X,N)$  and  $p(T2|A,Z,T1,X,N)$ 

repeat{
  ## sample  $T1/(1+T1)$  and  $T2/(1+T2)$  from beta dists
  T1.trans<-rbeta(1,shape1=sum(X*Z)+0.5,shape2=sum(Z)+0.5)
  T2.trans<-rbeta(1,shape1=sum(X)-sum(X*Z)+0.5,shape2=
    N[i-1]-sum(Z)+0.5)

  ## back transform to  $T1$  and  $T2$ 
  T1[i]<-T1.trans/(1-T1.trans)
  T2[i]<-T2.trans/(1-T2.trans)

  ## keep  $T1 < T2$ 
  if(T1[i]<T2[i])
    break
}

## sample from  $p(N|A,T1,T2,Z,X)$ 

## make sure  $N.new \geq w.tau$ 
repeat {
  N.new<-rnbino(1,mu=N[i-1],size=Metropolis.stdev.N)
  if(N.new>w.tau-1)
    break
}

## calculate  $\log(N.new!/(N.new-w.tau)!)$ 
N3.new<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.new[j+1]<-log(N.new-j)
}
N2.new<-sum(N3.new)

## calculate  $\log(N[i-1]!/(N[i-1]-w.tau)!)$ 
N3<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3[j+1]<-log(N[i-1]-j)
}
N2<-sum(N3)

## calculate log of acceptance ratio
logr1<-(-1/2)*log(N.new)+N2.new+N.new*log(A[i]*(1/(1+
  T1[i]))+(1-A[i])*(1/(1+T2[i])))+(1/2)*log(N[i-1])-N2-

```

```

      N[i-1]*log(A[i]*(1/(1+T1[i]))+(1-A[i])*(1/(1+T2[i])))
      +log(dnbinom(N[i-1],mu=N.new,size=Metropolis.stdev.N))
      -log(dnbinom(N.new,mu=N[i-1],size=Metropolis.stdev.N))

## calculate acceptance ratio
r1<-exp(logr1)

## accept or reject the proposed value
if (runif(1)<min(r1,1)) {N[i]<-N.new ; a1<-a1+1}
  else N[i]<-N[i-1]

## calculate deviance from current sample

## calculate log(N[i]/(N[i]-w.tau)!)
N3.curr<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.curr[j+1]<-log(N[i]-j)
}
N2.curr<-sum(N3.curr)

# calculate deviance
D.post[i]<-(-2)*(N2.curr+(N[i]-w.tau)*log(A[i]*(1/(1+
  T1[i]))+(1-A[i])*(1/(1+T2[i]))) +sum(freqdata[,2]
  *log(A[i]*(1/(1+T1[i]))*(T1[i]/(1+T1[i]))^freqdata[,1]
  +(1-A[i])*(1/(1+T2[i]))*(T2[i]/(1+T2[i]))^freqdata[,1]))
  -sum(log(factorial(freqdata[,2]))))
}

```

Step 4: model diagnostics

```

## 1) deviance at posterior mean
mean.A<-mean(A[(burn.in+1):iterations])
mean.T1<-mean(T1[(burn.in+1):iterations])
mean.T2<-mean(T2[(burn.in+1):iterations])
mean.N<-mean(N[(burn.in+1):iterations])

## calculate log(mean.N/(mean.N-w.tau)!)
N3.mean<-rep(0,w.tau)
for (j in 0:(w.tau-1)){
  N3.mean[j+1]<-log(mean.N-j)
}

```

```

    }
    N2.mean<-sum(N3.mean)

D.mean<-(-2)*(N2.curr+(mean.N-w.tau)*log(mean.A*(1/(1+mean.T1))
+(1-mean.A)*(1/(1+mean.T2))))+sum(freqdata[,2]*log(mean.A
*(1/(1+mean.T1))*(mean.T1/(1+mean.T1))^freqdata[,1]+(1
-mean.A)*(1/(1+mean.T2))*(mean.T2/(1+mean.T2))
^freqdata[,1]))-sum(log(factorial(freqdata[,2]))))

## 2) posterior mean and median deviances
mean.D<-mean(D.post[(burn.in+1):iterations])
median.D<-quantile(D.post[(burn.in+1):iterations],probs=.5,
names=F)

## 3) model complexity
p.D<-mean.D-D.mean

## 4) Deviance information criterion
DIC<-2*mean.D-D.mean

### Step 5: fitted values based on medians of the marginal
posterior
median.A<-quantile(A[(burn.in+1):iterations],probs=.5,
names=F)
median.T1<-quantile(T1[(burn.in+1):iterations],probs=.5,
names=F)
median.T2<-quantile(T2[(burn.in+1):iterations],probs=.5,
names=F)
median.N<-quantile(N[(burn.in+1):iterations],probs=.5,
names=F)

fits<-rep(0,tau)
for (k in 1:tau){
  fits[k]<-(median.N)*(median.A*dgeom(k,prob=1/(1
+median.T1)))+(1-median.A)*dgeom(k,prob=1/(1+median.T2)))
}
fitted.values<-data.frame(cbind(j=seq(1,tau),fits,count=
freqdata[,2]))

```

```

### Step 6: results

# this library is needed to use function export.eps
library(HH)
hist.points<-hist(N[(burn.in+1):iterations]+w-w.tau,breaks=
  seq(w,max(N)+w-w.tau+1)-0.5)

results<-data.frame(
  filename=filename,
  w=w,
  n=n,
  NP.est.C=NP.est.n0+w,
  tau=tau,
  w.tau=w.tau,
  n.tau=n.tau,
  iterations=iterations,
  burn.in=burn.in,
  acceptance.rate.N=a1/iterations,
  mode.N=hist.points$mids[which.max(hist.points$intensities)],
  mean.N=mean(N[(burn.in+1):iterations])+w-w.tau,
  median.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.5,
    names=F),
  LCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.025,
    names=F),
  UCI.N=quantile(N[(burn.in+1):iterations]+w-w.tau,probs=.975,
    names=F),
  mean.D=mean.D,
  median.D=median.D,
  DIC
)

# print results and fitted values
print(t(results),quote=F)
print(fitted.values)

# output results and fitted values
write.csv(t(results),filename.analysis,quote=F)
write.csv(fitted.values,filename.fits,quote=F)

# trace plot for C

# first thin values of C if there are more than 10,000
iterations

```

```

# must be a divisor of (iterations-burn.in)
iterations.trace<-min(10000,iterations-burn.in)
N.thin<-rep(0,iterations.trace)
for (k in 1:iterations.trace){
N.thin[k]<-N[k*((iterations-burn.in)/iterations.trace)]
}

# make trace plot
plot(1:iterations.trace,N.thin,xlab="",ylab="")
mtext("Iteration Number",1,line=2.5,cex=1.5)
mtext("Total Number of Species",2,line=2.5,cex=1.5)
export.eps(filename.trace)

# autocorrelation plot for C
my2.acf<-acf(N[(burn.in+1):iterations],type="correlation")
plot(my2.acf,main="",ylab="",xlab="")
  mtext("Lag",1,line=2.5,cex=1.5)
  mtext("ACF",2,line=2.5,cex=1.5)
export.eps(filename.AC)

# histogram of C with a bar for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1)-0.5,main="",xlab="",
  col='purple',freq=F,ylab="")
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.C)

# a histogram with # bars for each discrete value
hist(N[(burn.in+1):iterations]+w-w.tau,breaks=seq(w,max(N[
  (burn.in+1):iterations])+w-w.tau+1,length=(max(N[
  (burn.in+1):iterations])-w.tau+1)/bars+1)-0.5,main="",
  xlab="",col='purple',freq=F,ylab="",xlim=c(w,quantile(N[
  (burn.in+1):iterations]+w-w.tau,probs=.975,names=F)))
mtext("Total Number of Species",1,line=2.5,cex=1.5)
box()
export.eps(filename.Cbars)

```

REFERENCES

1. Behnke, A., Bunge, J., Barger, K., Breiner, H-W., Alla, V. and T. Stoeck (2006) Microeukaryote Community Patterns along an O₂/H₂S Gradient in a Supersulfidic Anoxic Fjord (Framvaren, Norway). *Applied and Environmental Microbiology*, **72**(5): 3626-3636.
2. Berger, J. O., Liseo, B. and L. Wolpert (1999) Integrated Likelihood Methods for Eliminating Nuisance Parameters. *Statistical Science*, **14**(1): 1-28.
3. Bernardo, J. M. (1979) Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, **41**: 113-147 (with discussion).
4. Bernardo, J. M. and J. M. Ramón (1998) An Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters. *The Statistician*, **47**(Part 1): 101-135.
5. Bernardo, J. M. and F. M. A. Smith (2000) *Bayesian Theory*. John Wiley & Sons, Ltd., Chichester.
6. Blumenthal (1977) Estimating Population Size with Truncated Sampling. *Communications in Statistics-Theory and Methods*, **A6**(4): 297-308.
7. Blumenthal (1992) Population Size Estimation with Truncated Data: A Survey. *Bayesian Analysis in Statistics and Econometrics*, **75**: 361-367.
8. Blumenthal, S., Dahiya, R. C., and A. J. Gross (1978) Estimating the Complete Sample Size from an Incomplete Poisson Sample. *Journal of the American Statistical Association*, **73**(361): 182-187.

9. Blumenthal, S., and R. Marcus (1975) Estimating Population Size with Exponential Failure. *Journal of the American Statistical Association*, **70**(352): 913-921.
10. Boender, C. G. E. and A. H. G. Rinnooy Kan (1987) A Multinomial Bayesian Approach to the Estimation of Population and Vocabulary Size. *Biometrika*, **74**(4): 849-856.
11. Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W. and C. Viwatwongkasem (2004) Estimating the Number of Drug Users in Bangkok 2001: A Capturerecapture Approach Using Repeated Entries in One List. *European Journal of Epidemiology*, **19**: 1075-1083.
12. Böhning, D. and D. Schön (2005) Nonparametric Maximum Likelihood Estimation of Population Size Based on the Counting Distribution. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, **54**: 721-737.
13. Booth, Andrew D. (1967) A 'Law' of Occurrences for Words of Low Frequency. *Information and Control*, **10**: 386-393.
14. Borgella, R. and T. A. Gavin (2005) Avian Community Dynamics in a Fragmented Tropical Landscape. *Ecological Application*, **15**(3): 1062-1073.
15. Brookmeyer, R. and M. H. Gail (1988) A Method For Obtaining Short-Term Projections and Lower Bounds on the Size of the Aids Epidemic. *Journal Of The American Statistical Association*, **83**(402): 301-308.
16. Buckland, S. T., Goudie, I. and D. Borchers (2000) Wildlife Population Assessment: Past Developments and Future Directions. *Biometrics*, **56**: 1-12.

17. Bunge, J. and M. Fitzpatrick (1993). Estimating the Number of Species: A Review. *Journal of the American Statistical Association*, **88**(421): 364-373.
18. Bunge, J. (accessed 4 Feb 2008) Estimating the Number of Classes in a Population: Bibliography. <<http://www.stat.cornell.edu/~bunge/bibliography.htm>>.
19. Burrell, Q. L. (1988) A Simple Empirical Method for Predicting Library Circulations. *Journal of Documentation*, **44**(4): 302-314.
20. Chao, A. (1987) Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. *Biometrics*, **43**(4): 783-791.
21. Cobabe, E. A. and W. D. Allmon (1994) Effects of Sampling on Paleoecological and Taphonomic Analyses in High-Diversity Fossil Accumulations-an Example from the Eocene Gosport Sand, Alabama. *Lethaia*, **27**(2): 167-178.
22. Cormack, R. M. (1992) Interval Estimation for Mark-Recapture Studies of Closed Populations. *Biometrics*, **48**: 567-576.
23. Ding, Ye (1996) On the Asymptotic Normality of Multinomial Population Size Estimators with Application to Backcalculation of AIDS Epidemic. *Biometrika*, **83**(3): 695-699.
24. Efron, B. and R. Tibshirani (1976) Estimating the Number of Unseen Species: How Many Words Did Shakespeare know? *Biometrika*, **63**: 435-447.
25. Esty, W. W. (1986) Estimation of the Size of a Coinage: A Survey and Comparison of Methods. *Numismatic Chronicle*, **146**: 185-215.

26. Fisher, R. A., A. S. Corbet and C. B. Williams (1943) The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. *Journal of Animal Ecology*, **12**: 42-58.
27. Gelman, A., Carlin, J. B., Stern, H. S. and D. B. Rubin (2004) *Bayesian Data Analysis*. Second edition, Chapman & Hall/CRC Press, Boca Raton.
28. George, E. I. and C. P. Robert (1992) Capture-recapture Estimation Via Gibbs Sampling. *Biometrika*, **79**(4): 677-683.
29. Good, I. J. (1953) The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*, **40**(3/4): 237-264.
30. Hill, B. M. (1979) Posterior Moments of the Number of Species in a Finite Population and the Posterior Probability of Finding a New Species. *Journal of the American Statistical Association*, **74**(367): 668- 673.
31. Hong, S-H., Bunge, J., Jeon, S-O. and S. S. Epstein (2006) Predicting Microbial Species Richness. *Proceedings of the National Academy of Sciences*, **103**(1): 117-122.
32. Hsieh, Y-H., Chen, C. W. S., Lee, S-M., Chen, Y-M. A., Wu, S-I., Lai, S-F. and A-L. Chang (2006) Estimating the Number of HIV-infected Gay Sauna Patrons in Taipei Area. *Physica A*, **362**: 495503.
33. Huber, J., Butterfield, D. and J. Baross (2003) Bacterial Diversity in a Subseafloor Habitat Following a Deep-sea Volcanic Eruption. *FEMS Microbiology Ecology*, **43**(3): 393-409.
34. Huber, J. A., Johnson, H. P., Butterfield, D. A. & J. A. Baross (2006) Microbial Life in Ridge Flank Crustal Fluids. *Environmental Microbiology*, **8**(1): 88-99.

35. Huzurbazar, V. S. (1950) Probability Distributions and Orthogonal Parameters. *Mathematical Proceedings of the Cambridge Philosophical Society*, **46**: 281-284.
36. Irony, T. Z. and N. D. Singpurwalla (1997) Non-informative Priors do not Exist: A Dialogue with José M. Bernardo. *Journal of Statistical Planning and Inference*, **65**: 159-189.
37. Jeffreys, H. (1939/1961). *Theory of Probability*. University Press, Oxford. Third edition in 1961, University Press, Oxford.
38. Jeffreys, H. (1946) An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **186**: 453-461.
39. Jeon, S-O., Bunge, J., Stoeck, T., Barger, K., Hong, S-H. and S. Epstein (2006) Synthetic Statistical Approach Reveals Spectacular Richness of Microbial Eukaryotes in Anoxic Water Column. *Applied and Environmental Microbiology*, **72**(10): 6578-6583.
40. Kass, R. E., and L. Wasserman (1996) The Selection of Prior Distributions by Formal Rules. *Journal of the American Statistical Association*, **91**(435): 1343-1370.
41. Kormas, K. A., Smith, D. C., Edgcomb, V. and A. Teske (2003) Molecular Analysis of Deep Subsurface Microbial Communities in Nankai Trough Sediments (ODP Leg 190, Site 1176). *FEMS Microbiology Ecology*, **45**(2): 115-125.
42. Kormas, K., Tivey, M., Von Damm, K. and A. Teske (2006) Bacterial and Archaeal Phylotypes Associated with Distinct Mineralogical Layers of a

- White Smoker Spire from a Deep-sea Hydrothermal Vent Site (9 N, East Pacific Rise). *Environmental Microbiology*, **8**(5): 909-920.
43. Lewins, W. A. and D. N. Joanes (1984) Bayesian Estimation of the Number of Species. *Biometrics*, **40**: 323-328.
 44. Ley, R. E., Harris, J. K., Wilcox, J., Spear, J. R., Miller, S. R., Bebout, B. M., Maresca, J. A., Bryant, D. A., Sogin, M. L. and N. R. Pace (2006) Unexpected Diversity and Complexity of the Guerrero Negro Hypersaline Microbial Mat. *Applied and Environmental Microbiology*, **72**(5): 3685-3695.
 45. Lindley, D. V. (1956) On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, **27**(4): 986-1005.
 46. Lindsay, B. G. and Roeder, K. (1987) A Unified Treatment of Integer Parameter Models. *Journal of the American Statistical Association*, **82**(399): 758-764.
 47. Lloyd, C. J., Yip, P. S. F. and K. S. Chan (1999) Estimating the Number of Faults: Efficiency of Removal, Recapture, and Seeding. *IEEE Transactions on Reliability*, **48**(4): 369-376.
 48. Madigan, D. and J. C. York (1997) Bayesian Methods for Estimation of the Size of a Closed Population. *Biometrika*, **84**(1): 19-31.
 49. Norris III, J. L. and Pollock, K. H. (1998) Non-parametric MLE for Poisson species abundance models allowing for heterogeneity between species. *Environmental and Ecological Statistics*, **5**: 391-402.
 50. Pollock, K. H. (2000) Capture-recapture Models. *Journal of the American Statistical Association*, **95**: 293-296.

51. Raftery, A. E. (1987) Inference and Prediction for a General Order Statistic Model With Unknown Population Size. *Journal of the American Statistical Association*, **82**(400): 1163-1168.
52. Raftery, A. E., and J. E. Zeh (1998) Estimating Bowhead Whale Population Size and Rate of Increase from the 1993 Census. *Journal of the American Statistical Association*, **93**(442): 451-463.
53. Rissanen, J. (1983) A Universal Prior for Integers and Estimation by Minimum Description Length. *The Annals of Statistics*, **11**(2): 416-431.
54. Rodrigues, J., Milan, L. A. and J. G. Leite (2001) Hierarchical Bayesian Estimation for the Number of Species. *Biometrical Journal*, **43**(6): 737-746.
55. Sanathanan, L. (1972) Estimating the Size of a Multinomial Population. *The Annals of Mathematical Statistics*, **43**(1): 142-152.
56. Sandland, R. L. and R. M. Cormack (1984) Statistical Inference for Poisson and Multinomial Models for Capture-Recapture Experiments. *Biometrika*, **71**(1): 27-33.
57. Santelli, C. M., Orcutt, B. N., Banning, E., Bach, W., Moyer, C. L., Sogin, M. L., Staudigel, H. and K. J. Edwards (29 May 2008) Abundance and Diversity of Microbial Life in Ocean Crust. *Nature*, **453**: 653-656.
58. Schwarz, C. and G. Seber (1999) Estimating Animal Abundance: Review iii. *Statistical Science*, **14**: 427-456.
59. Seber, G. A. F. and C. J. Schwarz (2000) Capture-recapture: Before and After EURING 2000. *Journal of Applied Statistics*, **29**: 5-18.

60. Smith, P. J. (1991) Bayesian Analyses for a Multiple Capture-Recapture Model. *Biometrika*, **78**(2): 399-407.
61. Smith, K. L. and M. L. Jones (2005) Watershed-Level Sampling Effort Requirements for Determining Riverine Fish Species Composition. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**(7): 1580-1588.
62. Solow, A. R. (1994) On the Bayesian Estimation of the Number of Species in a Community. *Ecology*, **75**(7): 2139-2142.
63. Stam, A. J. (1987) Statistical Problem in Ancient Numismatics. *Statistica Neerlandica*, **41**(3): 151-173.
64. Tanner, M. A. and W. H. Wong (1987) The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, **82**(398): 528-540.
65. Tardella, L. (2002) A New Bayesian Method for Nonparametric Capture-recapture Models in Presence of Heterogeneity. *Biometrika*, **89**(4): 807-817.
66. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y-H. and H. O. Smith (2004) Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*, **304**(5667): 66-74.
67. Walther, B. A. and J-L. Martin (2001) Species Richness Estimation of Bird Communities: How to Control for Sampling Effort? *Ibis*, **143**(3): 413-419.

68. Wang, X., He, C. Z., and Sun, D. (2007) Bayesian Population Estimation for Small Sample Capture-recapture Data Using Noninformative Priors. *Journal of Statistical Planning and Inference*, **137**(4): 1099-1118.
69. Wang, Ji-Ping Z. and Bruce G. Lindsay (2005) A Penalized Nonparametric Maximum Likelihood Approach to Species Richness Estimation. *Journal of the American Statistical Association*, **100**(471): 942-959.
70. Williams, C. B. (1939) An Analysis of Four Years Captures of Insects in a Light Trap. Part I. General survey; Sex Proportion; Phenology; and Time of Flight. *Transactions of the Royal Entomological Society of London*, **89**: 79-131.
71. Zeh, J. E., Ko, D., Krogman, B. D. and R. Sonntag (1986) A Multinomial Model for Estimating the Size of a Whale Population from Incomplete Census Data. *Biometrics*, **42**(1): 1-14.
72. Zuendorf, A., Behnke, A., Bunge, J., Barger, K., and T. Stoeck (2006) Diversity Estimates of Microeukaryotes Below the Chemocline of the Anoxic Mariager Fjord, Denmark. *FEMS Microbiology Ecology*, **58**(3): 476-491.